IDC Engineering Pocket Guide

Volume 6
Industrial Automation

Rev 2.2

Edited by
Srinivas Medida

Technology Training that Works

IDC Technologies Pty Ltd
PO Box 1093, West Perth, Western Australia 6872
Offices in Australia, New Zealand, Singapore, United Kingdom, Ireland, Malaysia, Poland, United States of America, Canada, South Africa, Vietnam and India

First published 2008

Disclaimer

Trademarks

Acknowledgements

IDC Technologies expresses its sincere thanks to all those engineers and technicians on our training workshops who freely made available their expertise in preparing this manual.

Contents

Preface

Industrial Automation is a discipline that includes knowledge and expertise from various branches of engineering including electrical, electronics, chemical, mechanical, communications and more recently computer and software engineering. Automation & Control by its very nature demands a cross fertilization of these faculties.

Industrial Automation Engineers have always drawn new technologies and implemented original or enhanced versions to meet their requirements. As the range of technology diversifies the demand on the innovative ability of these Engineers has increased.

IDC Technologies has been in the business of bringing together the domain gurus and the practicing engineers under an umbrella called training. The sum of the knowledge that IDC Technologies has acquired over many years has now given it an opportunity to compile this comprehensive hand book for the reference of every automation engineer.

The breadth and depth of Industrial Automation is enormous and justice cannot be expected from a book of a few hundred pages. This book comprises over 1200 pages of useful, hard hitting information from the trenches on industrial automation. This book delivers a critical blend of knowledge and skills, covering technology in control and instrumentation, industry analysis and forecasts, leadership and management - everything that is relevant to a modern control and instrumentation engineer. Good management, financial and business skills are also provided in these chapters. These highly practical materials provide you with solid skills in this often neglected area for control and instrumentation engineers.

This book was originally written for UK and other European users and contains many references to the products and standards in those countries. We have made an effort to include IEEE/ANSI/NEMA references wherever possible. The general protection approach and theoretical principles are however universally applicable.

The terms '*earth*' as well as '*ground*' have both been in general use to describe the common power/signal reference point interchangeably around the world in the Electro-technical terminology. While the USA and other North American countries favor the use of the term *'ground'*, European countries including the UK and many other Eastern countries prefer the term *'earth'*. In this book, we chose to adopt the term *'ground'* to denote the common electrical reference point. Our sincere apologies to those readers who would have preferred the use of the term '*earth*'.

Chapter 1. Introduction

Society in its daily endeavours has become so dependent on automation that it is difficult to imagine life without automation engineering. In addition to the industrial production with which it is popularly associated, it now covers a number of unexpected areas. Trade, environmental protection engineering, traffic engineering, agriculture, building engineering, and medical engineering are but some of the areas where automation is playing a prominent role. Automation engineering is a cross sectional discipline that requires proportional knowledge in hardware and software development and their applications. In the past, automation engineering was mainly understood as control engineering dealing with a number of electrical and electronic components. This picture has changed since computers and software have made their way into every component and element of communications and automation.

Industrial automation engineers carry a lot of responsibility in their profession. No other domain demands so much quality from so many perspectives of the function, yet with significant restrictions on the budget. The project managers of industrial automation projects have significant resource constraint, considering the ever changing demands of its management, trying to adopt the rapid acceleration of the technological changes and simultaneously trying to maintain the reliability and unbreakable security of the plant and its instruments.

This book is structured to walk you through a précised life cycle of the various automation activities of a plant. There are a number of books that cover different aspects of automation but this is all encompassing.

Chapter 2. I&C Drawings and Documentation

2.1. Introduction to Plant Design

Plant design (process plant design, power plant design, etc.) refers to the automation technologies, work practices and business rules supporting the design and engineering of process and power plants. Such plants can be built for chemical, petroleum, utility, shipbuilding, and other facilities. Plant design is used to designate a general market area by the many vendors offering technologies to support plant design work.

2.2. Process diagrams

The ‘process’ is an idea or concept that is developed to a certain level in order to determine the feasibility of the project. ‘Feasibility’ study is the name given to a small design project that is conducted to determine the scope and cost of implementing the project from concept to operation.

To keep things simple, for example, design an imaginary coffee bottling plant to produce bottled coffee for distribution. Start by creating a basic flow diagram that illustrates the objective for the proposed plant; this diagram is called a “Process Block Diagram”.

2.2.1. Process block diagram

The block diagram shown in Figure 2.1 is where it all starts. It is here that the basic components are looked at and the basic requirements determined. This is a diagram of the concept, giving a very broad view of the process.

The example below has all ingredients listed and shows that milk, sugar and black coffee make up different permutations of the final product. With this philosophy diagram complete, there is a need to determine the technical requirements. This is done by simultaneously developing two documents; the ‘Process Flow Diagram’ and the ‘Process Description Manual’.

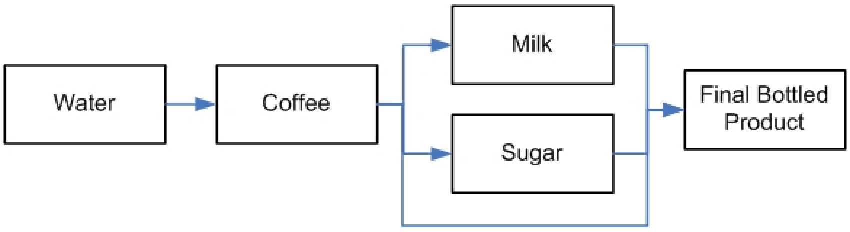


Figure 2.1
Basic flow diagram of Coffee bottling plant

2.2.2. Process flow diagram or piping flow diagram (PFD)

The PFD is where we start to define the process by adding equipment and the piping that joins the various items of equipment together. The idea behind the PFD is to show the entire process (the big picture) on as few drawing sheets as possible, as this document is used to develop the process plant and therefore the process engineer wants to see as much of the process as possible. This document is used to determine details like the tank sizes and pipe sizes.

Those familiar with mimic panels and SCADA flow screens will notice that these resemble the PFD more than the piping and instrumentation diagram (P&ID) with the addition of the instruments, but not the instrument function.

Mass balance: In its most simple form, what goes in must come out. The totals at the end of the process must equal the totals fed into the system.

2.2.3. Process description

The process description details the function / purpose of each item of equipment in the plant. This description should contain the following information:

- Installation operation – The installation produces bottled coffee
- Operating principles – Each part of the process is described
- Water supply – Filtered water at ambient temperature is supplied to the water holding tank, the capacity of the tank should be sufficient for all recipes
- Coffee supply – Due to the viscosity of the coffee syrup, the syrup is fed from a pressurized vessel to the autoclave, this line should be cleaned frequently with warm water. There will be batches of caffeinated and decaffeinated coffee, the coffee tanks and pipelines must be thoroughly cleaned between batches
- Milk supply – There will be an option for low fat or full cream milk, the milk supply should be sufficient for three days operation and should be kept as close to freezing as possible to ensure longevity of the milk
- Sugar supply – Sugar will be supplied in a syrup form, we will offer the coffee with no sugar, 1 teaspoon (5 ml of syrup) or two teaspoons (10 ml of syrup). Syrup lines must be cleaned on a regular basis
- Circuit draining/make-up – How to start-up or shutdown the facility, cleaning and flushing

- Liquid characteristics – A detailed description on analysis of each liquid type in the system. Includes specific gravity, viscosity, temperature, pressure, composition etc.
- Specific operating conditions linked to the process – The installation operates 24 hours a day, 365 days a year. As the installation deals with foodstuff, all piping and vessels are to be manufactured from stainless steel
- Specific maintenance conditions linked to the process – Hygiene levels to be observed
- Specific safety conditions linked to the process – Hygiene, contamination of product
- Performance requirements – This section describes the amount of product the plant must be able to produce in a given time frame.

PFD now starts to look something like the Figure 2.2 shown below.

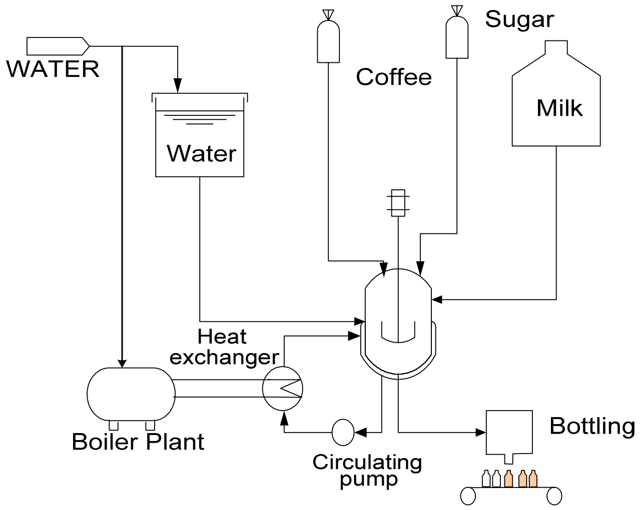


Figure 2.2
Process flow diagram

2.2.4. Piping and Instrumentation Diagram (P&ID)

The Piping & Instrumentation Diagram, which may also be referred to as the Process & Instrumentation Diagram, gives a graphical representation of the process including hardware (Piping, Equipment) and software (Control systems); this information is used for the design construction and operation of the facility.

The PFD defines "The flow of the process" The PFD covers batching, quantities, output, and composition.

The P&ID also provides important information needed by the constructor and manufacturer to develop the other construction input documents (the isometric drawings, or orthographic physical layout drawings, etc.). The P&ID provides direct input to the field for the physical design and installation of field-run piping. For clarity, it is usual to use the same general layout of flow paths on the P&ID as used in the flow diagram.

The P&ID ties together the system description, the flow diagram, the electrical control schematic, and the control logic diagram. It accomplishes this by showing all of the piping, equipment, principal instruments, instrument loops, and control interlocks. The P&ID contains a minimum of text in the form of notes (the system description minimizes the need for text on the P&ID).

The typical plant operation's environment uses the P&ID as the principal document to locate information about the facility, whether this is physical data about an object, or information, such as financial, regulatory compliance, safety, HAZOP information, etc.

The P&ID defines "The control of the flow of the process" where the PFD is the main circuit; the P&ID is the control circuit. Once thoroughly conversant with the PFD & Process description, the engineers from the relevant disciplines (piping, electrical & control systems) attend a number of HAZOP sessions to develop the P&ID.

2.2.5. P&ID standards

Before development of the P&ID can begin, a thorough set of standards is required. These standards must define the format of each component of the P&ID. The following should be shown on the P&ID:

- Mechanical Equipment
- Equipment Numbering
- Presentation on the P&ID
- Valves
- Hand valves
- Control valves
- Piping
- Pipe numbering
- Nozzles & Flanges
- Equipment & instrument numbering systems

A completed P&ID may therefore appear as shown in Figure 2.3.

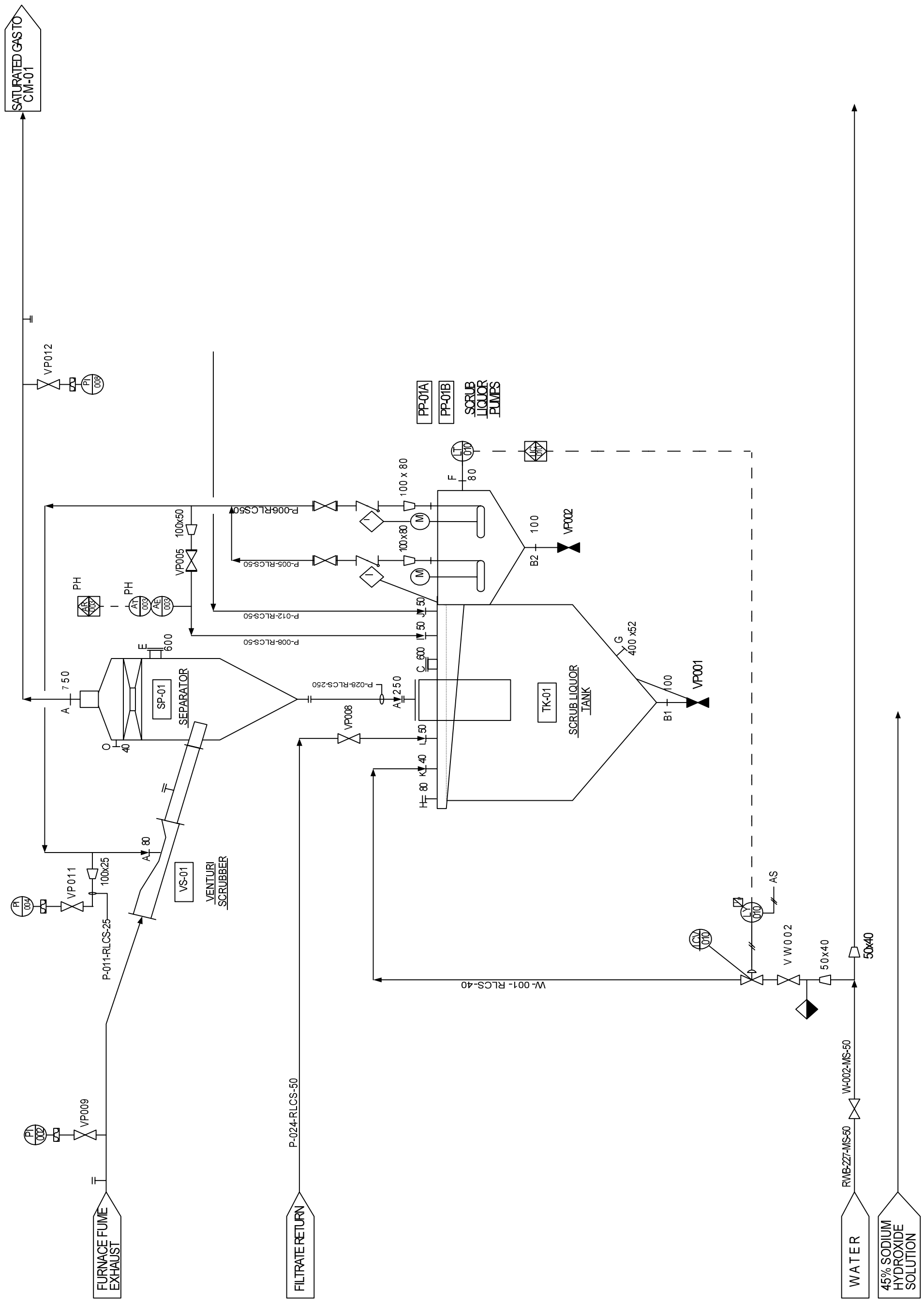


Figure 2.3
Completed P&ID

2.3. Instrumentation documentation

Instrumentation documentation consists of drawings, diagrams and schedules. The documentation is used by various people for different purposes. Of all the disciplines in a project, instrumentation is the most interlinked and therefore the most difficult to control.

The best way to understand the purpose and function of each document is to look at the complete project flow from design through to commissioning.

- Design
- Design criteria, standards, specifications, vendor lists
- Construction
- Quantity surveying, disputes, installation contractor, price per meter, per installation
- Operations
- Maintenance commissioning

2.3.1. Instrument list

This is a list of all the instruments on the plant, in the 'List' format. All the instruments of the same type (tag) are listed together; for example, all the pressure transmitters 'PT' are grouped together.

Table 2.1
Instrument list

Instrument index lists	Associated documentation such as loop drawing number, datasheets, installation details and P&ID.
Loop List	The same information as the instrument list but ordered by loop number instead of tag number. This sort of order will group all elements of the same loop number together.
Function	Gives a list of all the instrumentation on the plant and may include 'virtual' instruments such as controllers in a DCS or PLC.
Tag No	The instrument tag number as defined by the specification.
Description	Description of the instrument as denoted by the tag number.
Service Description	A description of the process related parameter.
Functional Description	The role of the device.
Manufacturer	Details of the manufacturer of the device.
Model	Details of the model type and number.

2.3.2. Instrument location plans

The instrument location drawing is used to indicate an approximate location of the instruments and junction boxes. This drawing is then used to determine the cable lengths from the instrument to the junction box or control room. This drawing is also used to give the installation contractor an idea as to where the instrument should be installed.

2.3.3. Cable racking layout

Use of the racking layout drawing has grown with the use of 3D CAD packages; this drawing shows the physical layout and sizes of the rack as it moves through the plant.

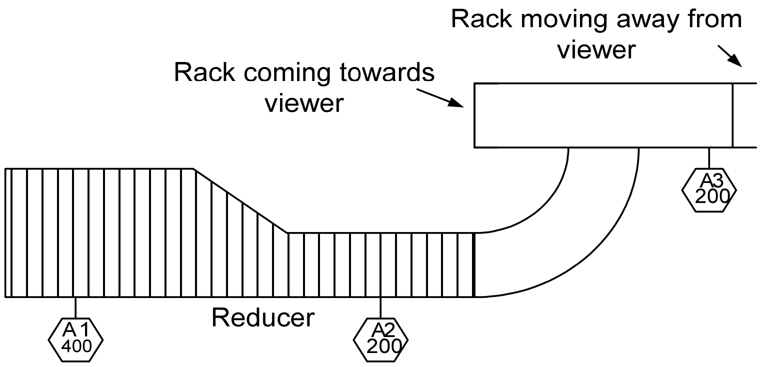


Figure 2.4
Cable racking layout

2.3.4. Cable routing layout

Prior to the advent of 3D CAD packages, the routing layout used a single line to indicate the rack direction as well as routing and sizes and was known as a 'Racking & Routing layout'.

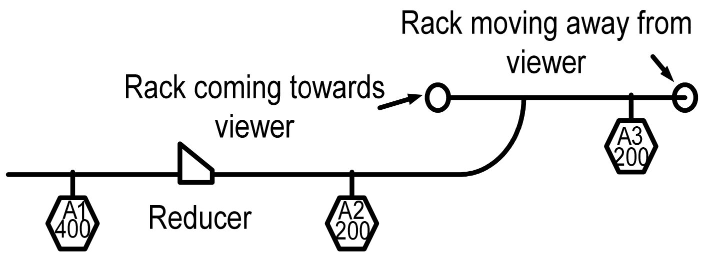


Figure 2.5
Cable routing layout

2.3.5. Block diagrams – signal, cable and power block diagrams

Cable block diagrams can be divided into two categories: Power and Signal block diagrams. The block diagram is used to give an overall graphical representation of the cabling philosophy for the plant.

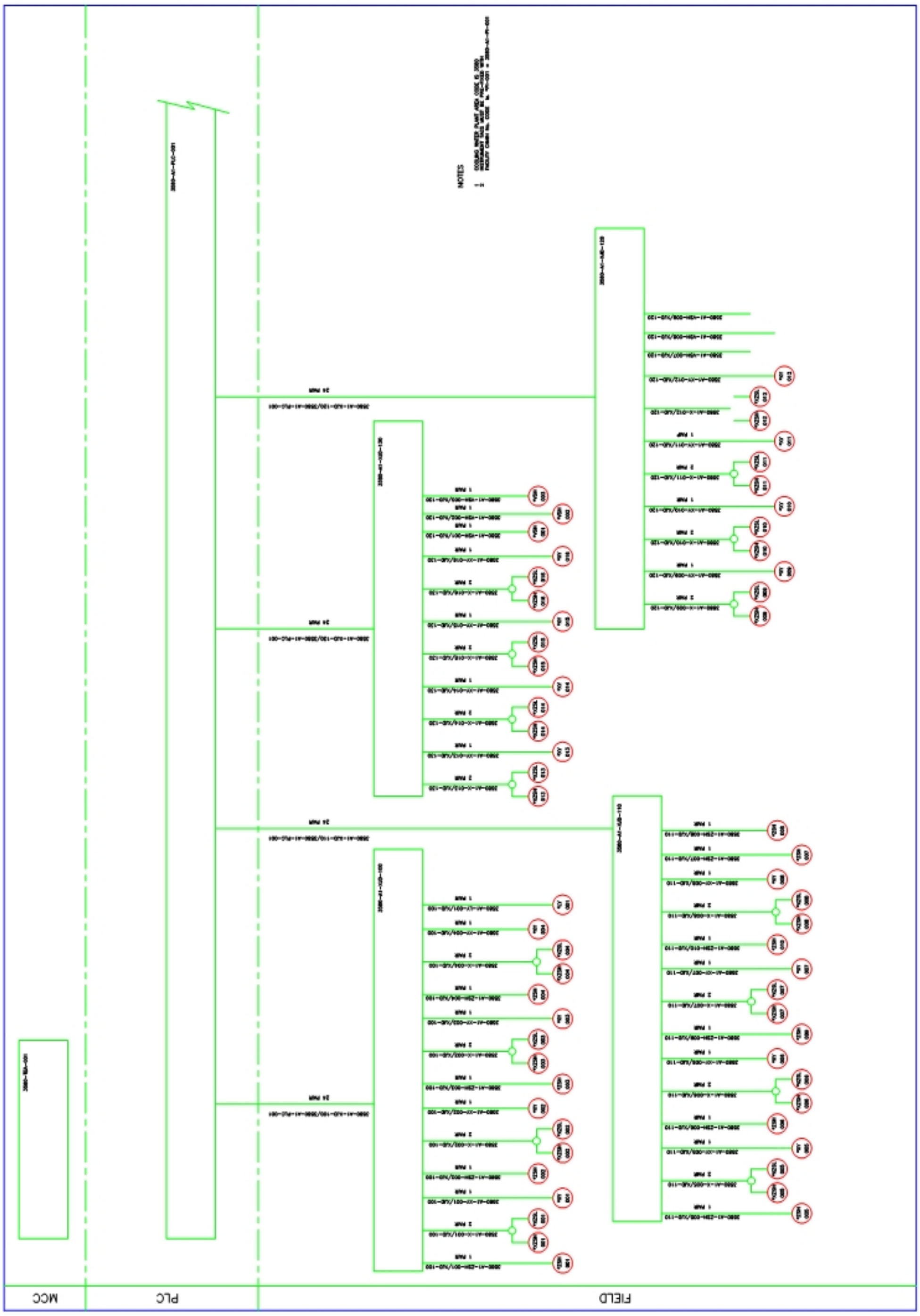


Figure 2.6
Block diagram

Table 2. 2
Field connections / Wiring diagrams

Field connections / Wiring diagrams	
Function	To instruct the wireman on how to wire the field cables at the junction box.
Used by	The installation contractor. When the cable is installed on the cable rack, it is left lying loose at both the instrument and junction box ends. The installation contractor stands at the junction box and strips each cable and wires it into the box according to the drawing.

Table 2. 3
Power distribution diagram

Power distribution diagram	
Function	There are various methods of supplying power to field instruments; the various formats of the power distribution diagrams show these different wiring systems.
Used by	Various people depending on the wiring philosophy, such as the panel wireman, field wiring contractor.

Table 2.4
Earthing diagram

Function	Used to indicate how the earthing should be done. Although this is often undertaken by the electrical discipline, there are occasions when the instrument designer may or must generate his own scheme – Eg. for earthing of zener barriers in a hazardous area environment.
Used by	Earthing contractor for the installation of the earthing. This drawing should also be kept for future modifications and reference.

Table 2.5
Loop diagrams

Function	A diagram that comprehensively details the wiring of the loop, showing every connection from field to instrument or I/O point of a DCS/PLC.
Used by	Maintenance staff during the operation of the plant and by commissioning staff at start up.

2.4. Electrical documentation

The electrical schematics section covers the layout of electrical schematic diagrams, lists and various symbols used.

2.4.1. The Load List

The load list is used to total the power supply requirements for each device per plant area or process. Load lists are made for each voltage level on the plant. The sample table shown below is a typical layout of a load list.

Table 2.6
Sample

Device	Voltage	Amps	Watts	VA	Total	Feeder
400-PMP-01	380					400-TAD-01

2.4.2. The Single Line Diagram

The single line diagram (sometimes called the one line diagram) uses single lines and standard symbols to show electrical cables, bus bars and component parts of a circuit or system of circuits. The single line diagram shows the overall strategy for system operation. Duplication of a 3-wire system is reduced by showing single devices on a single wire. These single line diagrams may be used in the monitoring and control systems like SCADA applications for the operation.

2.4.3. The schematic diagram (main and circuit)

Schematic diagram shows both the main circuit and the control circuit in far greater detail; here all three lines of a 3-phase system are shown. The schematic shows the detailed layout of the control circuit for maintenance and faultfinding purposes rather than the overall picture presented by the single line diagram.

A schematic diagram shows the following main features:

- Main circuits
- Control, signal and monitoring circuits
- Equipment identification symbols with component parts and connections
- Equipment and terminal numbering
- Cross references – indicating where on the diagram or sequential sheet, the related parts of the equipment can be found.

2.4.4. Plant layout drawings

The plant layout drawing gives a physical plant layout, where equipment is drawn to resemble the plant item it represents.

2.4.5. Racking and Routing

These drawings are used to show the layout of the plant racking systems, the size of the racks and the cable numbers of all the cables running on that section of the rack.

2.4.6. Installation Details

The installation detail shows the layout of the equipment and gives an itemized list of all the equipment on the drawing as well as notes on the installation.

2.4.7. Panel Layout

The panel layout drawing gives the dimensions of the panel, the layout of the equipment in the panel, an itemized list of all the equipment used as well as quantities. The notes detail various items like specification references (paint, powder coating) and general notes.

2.4.8. Other electrical documents

Cable schedule: This is used mainly for installation purposes. It gives a source and destination for each cable and specifies the type of cable.

Point to point schedule: This facilitates wiring installation by showing termination points at each end of every wire.

Hazardous area drawings: A plant location drawing (in both plan and elevation) which shows, by means of overlays, plant area classifications (by zone and gas group) for potential leak hazards throughout a plant.

Ladder Logic Schematics: These are detailed schematics of a ladder structure where the discrete rungs represent control circuits in an overall scheme. These are most often used in the basic IEC programming language in PLCs, but are sometimes used in hardwired relay circuits.

Chapter 3. Process control

3.1. Basic Control Concepts

Most basic process control systems consist of a control loop as shown in Figure 3.1. This has four main components which are:

- A measurement of the state or condition of a process
- A controller calculating an action based on this measured value against a pre-set or desired value (set point)
- An output signal resulting from the controller calculation which is used to manipulate the process action through some form of actuator
- The process itself reacting to this signal, and changing its state or condition.

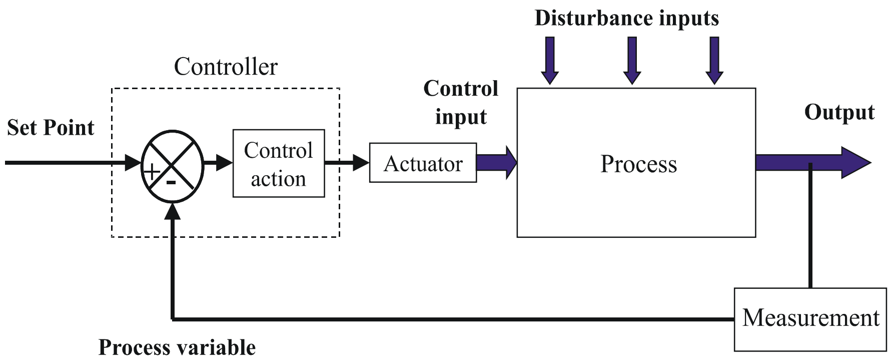


Figure 3.1
Block diagram showing the elements of a process control loop

Two of the most important signals used in process control are called

- Process Variable or PV
- Manipulated Variable or MV

In industrial process control, the Process Variable or PV is measured by an instrument in the field and acts as an input to an automatic controller which takes action based on the value of it. Alternatively, the PV can be an input to a data

display so that the operator can use the reading to adjust the process through manual control and supervision.

The variable to be manipulated, in order to have control over the PV, is called the Manipulated Variable. If we control a particular flow for instance, we manipulate a valve to control the flow. Here, the valve position is called the Manipulated Variable and the measured flow becomes the Process Variable.

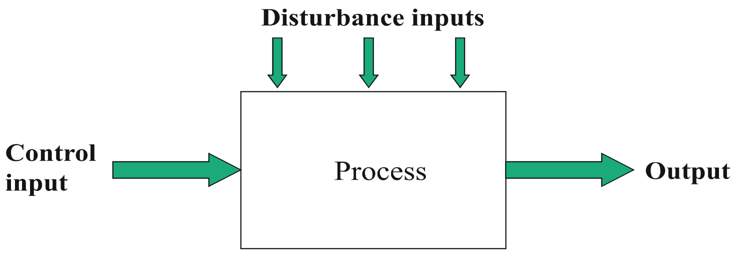
3.2. Principles of Control Systems

To perform an effective job of controlling a process, we need to know how the control input we are proposing to use will affect the output of the process. If we change the input conditions we need to know the following:

- Will the output rise or fall?
- How much response will we get?
- How long will it take for the output to change? .
- What will be the response curve or trajectory of the response?

The answers to these questions are best obtained by creating a mathematical model of the relationship between the chosen input and the output of the process in question. Process control designers use a very useful technique of block diagram modeling to assist in the representation of the process and its control system. The following section introduces the principles that should apply to most practical control loop situations.

The process plant is represented by an input/output block as shown in Figure 3.2.



Control inputs are also known as "manipulated variables"
The output is the process variable to be controlled

Figure 3.2
Basic block diagram for the process being controlled

In Figure 3.2, we see a controller signal that will operate on an input to the process, known as the 'manipulated variable'. We try to drive the output of the process to a particular value or set point by changing the input. The output may also be affected by other conditions in the process or by external actions such as changes in supply pressures or in the quality of materials being used in the process. These are all regarded as 'disturbance inputs' and our control action will need to overcome their influences as well as possible.

The challenge for the process control designer is to maintain the controlled process variable at the target value or change it to meet production needs whilst compensating for the disturbances that may arise from other inputs. So for

example, if we want to keep the level of water in a tank at a constant height while others are drawing off from it, we will manipulate the input flow to keep the level steady.

The value of a process model is that it provides a means of showing the way the output will respond to the input actions. This is done by having a mathematical model based on the physical and chemical laws affecting the process.

For example in Figure 3.3, an open tank with cross sectional area A is supplied with an inflow of water Q_1 that can be controlled or manipulated. The outflow from the tank passes through a valve with a resistance R to the output flow Q_2. The level of water or pressure head in the tank is denoted as H. We know that Q_2 will increase as H increases and when Q_2 equals Q_1 the level will become steady.

The block diagram of this process is shown in Figure 3.4.

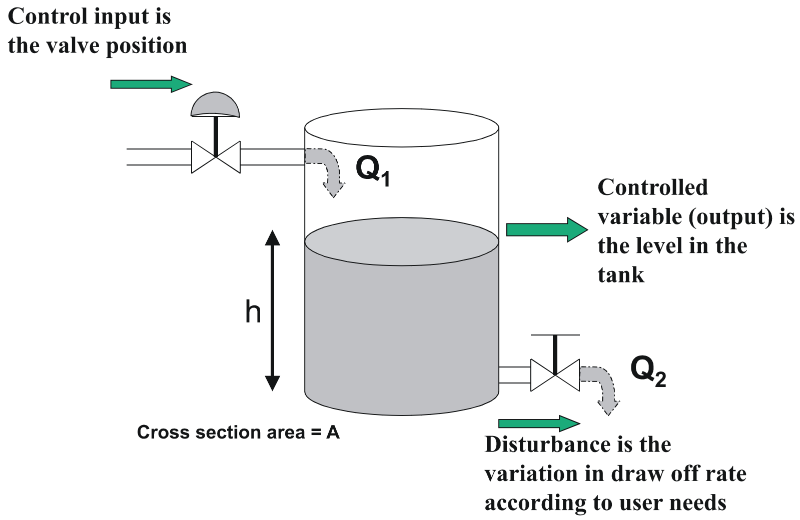


Figure 3.3
Example of a water tank with controlled inflow

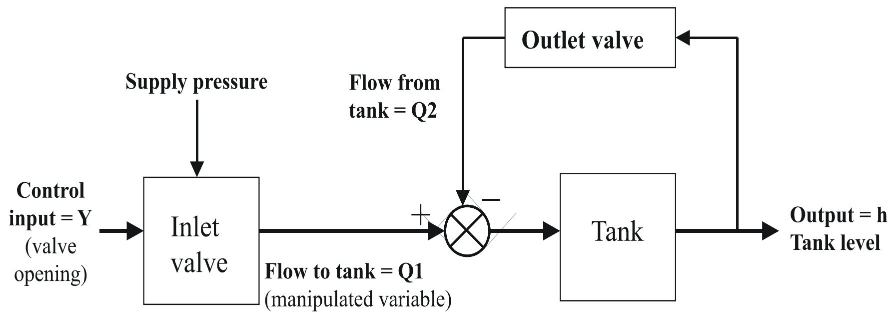


Figure 3.4
Elementary block diagram of tank process

3.2.1. Stability

A closed loop control system is stable if there is no continuous oscillation. A noisy and disturbed signal may show up as a varying trend; but it should never be confused with loop instability. The criteria for stability are these two conditions:

- The Loop Gain (KLOOP) for the critical frequency <1;
- Loop Phase Shift for the critical frequency < 180°.

3.2.2. Loop gain for critical frequency

Consider the situation where the total gain of the loop for a signal with that frequency has a total loop phase shift of 180°. A signal with this frequency is decaying in magnitude, if the gain for this signal is below 1. The other two alternatives are:

- Continuous oscillations which remain steady (Loop Gain = 1);
- Continuous oscillations which are increasing, or getting worse (Loop Gain > 1).

3.2.3. Loop phase shift for critical frequency

Consider the situation where the total phase shift for a signal with that frequency has a total loop gain of 1. A signal with this phase shift of 180° will generate oscillations if the loop gain is greater than 1. Increasing the Gain or Phase Shift destabilizes a closed loop, but makes it more responsive or sensitive.

Decreasing the Gain or Phase Shift stabilizes a closed loop at the expense of making it more sluggish.

The gain of the loop (K_{LOOP}) determines the OFFSET value of the controller; and offset varies with Set point changes.

3.2.4. Control Modes

There are five basic forms of control available in Process Control:

- On-Off
- Modulating
- Open Loop
- Feed Forward
- Closed loop

On-Off control: The oldest strategy for control is to use a switch giving simple on-off control, as illustrated in Figure 3.5. This is a discontinuous form of control action, and is also referred to as two-position control. A perfect on-off controller is 'on' when the measurement is below the set point (SP) and the manipulated variable (MV) is at its maximum value. Above the SP, the controller is 'off' and the MV is at a minimum.

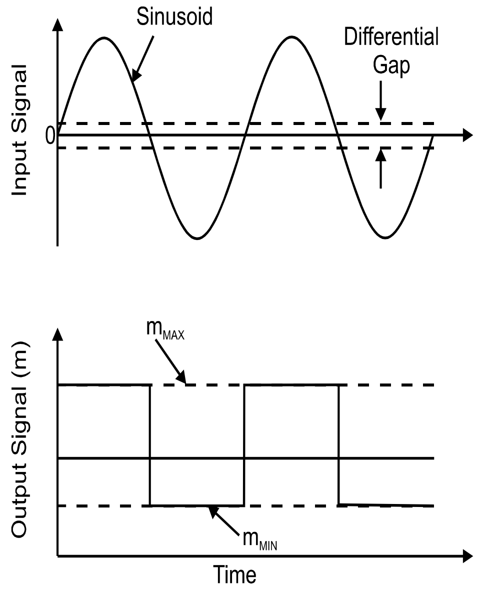


Figure 3.5
Response of a two positional controller to a sinusoidal input

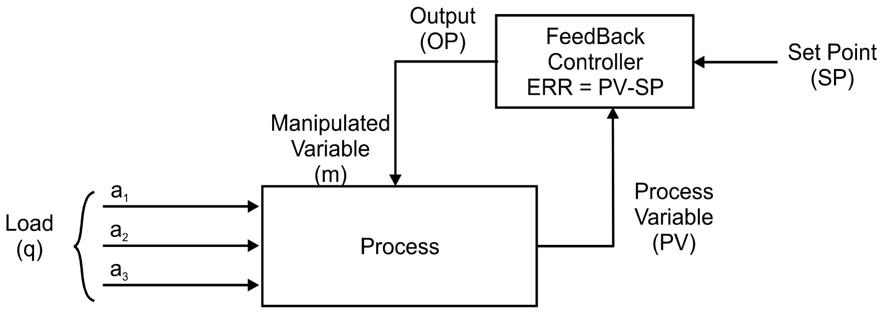
Modulating control: If the output of a controller can move through a range of values, this is modulating control.

Modulation Control takes place within a defined operating range only. That is, it must have upper and lower limits. Modulating control is a smoother form of control than step control. It can be used in both open loop and closed loop control systems.

Open loop control: Open loop control is thus called because the control action (Controller Output Signal OP) is not a function of the PV (Process Variable) or load changes. The open loop control does not self-correct, when these PV's drift.

Feed forward control: Feed forward control is a form of control based on anticipating the correct manipulated variables required to deliver the required output variable. It is seen as a form of open loop control as the PV is not used directly in the control action.

Closed loop or feedback control: If the PV, the objective of control, is used to determine the control action it is called closed loop control system. The principle is shown below in Figure 3.6.



$$OP = (m) = K_C \left(E_{RR} + \frac{1}{T_{int}} \int E_{RR}\, dt + T_{der} \frac{d}{dt} E_{RR} \right) + \text{Manual}$$

Figure 3.6
The feedback control loop

The idea of closed loop control is to measure the PV (Process Variable); compare this with the SP (Set Point), which is the desired, or target value; and determine a control action which results in a change of the OP (Output) value of an automatic controller.

In most cases, the ERROR (ERR) term is used to calculate the OP value.

ERR = PV - SP

If ERR = SP - PV has to be used, the controller has to be set for REVERSE control action.

3.3. Control modes in closed loop control

Most Closed loop Controllers are capable of controlling with three control modes which can be used separately or together

- Proportional Control (P)
- Integral, or Reset Control (I)
- Derivative, or Rate Control (D)

3.3.1. Proportional control(P)

This is the principal means of control. The automatic controller needs to correct the controllers OP, with an action proportional to ERR. The correction starts from an OP value at the beginning of automatic control action.

Proportional error and manual value: This is called as starting value manual. In the past, this has been referred to as "manual reset". In order to have an automatic correction made, that means correcting from the manual starting term, we always need a value of ERR. Without an ERR value there is no correction and go back to the value of manual.

Proportional band: Controllers Proportional Band is usually defined, in percentage terms, as the ratio of the input value, or PV to a full or 100% change in the controller output value or MV.

3.3.2. Integral control(I)

Integral action is used to control towards no OFFSET in the output signal. This means that it controls towards no error (ERR = 0). Integral control is normally used to assist proportional control. The combination of both is called as PI-control.

Formula for I-Control:

$$OP = \left(\frac{K}{T\ \text{int}}\right)\int_{O}^{T} \text{ERR dt}$$

Formula for PI-Control:

$$OP = \left(\frac{K}{T\ \text{int}}\right)\int_{O}^{T} \text{ERR dt} + (\text{K} * \text{ERR} + \text{MAUAL})$$

T_{int} is the Integral Time Constant.

3.3.3. Derivative control (D)

The only purpose of derivative control is to add stability to a closed loop control system. The magnitude of derivative control (D-Control) is proportional to the rate of change (or speed) of the PV.

Since the rate of change of noise can be large, using D-Control as a means of enhancing the stability of a control loop is done at the expense of amplifying noise. As D-Control on its own has no purpose, it is always used in combination with P-Control or PI-Control. This results in a PD-Control or PID-Control. PID-Control is mostly used if D-Control is required.

Formula for D-Control:

$$OP = K * Tder\ (dERR\ /dt)$$

Where

T_{der} is the Derivative Time Constant.

3.4. Tuning of Closed Loop Control

There are often many and sometimes contradictory objectives, when tuning a controller in a closed loop control system. The following list contains the most important objectives for tuning of a controller:

Minimization of the integral of the error : The objective here is to keep the area enclosed by the two curves, the SP and PV trends; to a minimum.

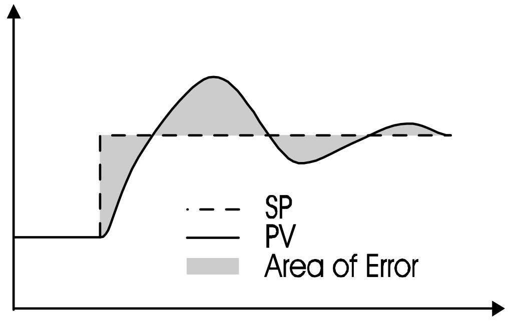


Figure 3.7
Integral on error

Minimization of the integral of the error squared: As shown in Figure 3.8, it is possible to have a small area of error but an unacceptable deviation of PV from SP for a start time. In such cases, special weight must be given to the magnitude of the deviation of PV from SP. Since the weight given is proportional to the magnitude of the deviation, the weight is multiplied by the error. This gives error squared (error squared = error * weight). Many modern controllers with automatic and continuous tuning work on this basis.

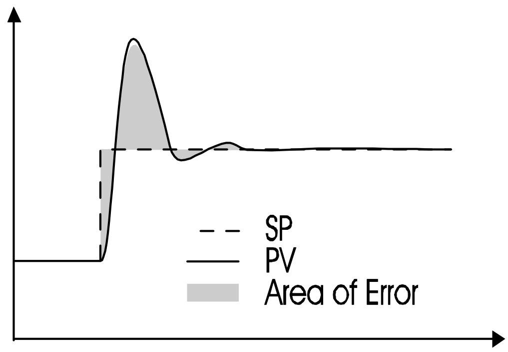


Figure 3.8
Integral on error square

Fast control: In most cases, fast control is a principle requirement from an operational point of view. However, this is principally achieved by operating the controller with a high gain. This quite often results in instability, or prolonged settling times from the effects of process disturbances.

Minimum wear and tear of controlled equipment: A valve or servo system for instance should not be moved unnecessarily frequently, fast or into extreme positions. In particular, the effects of noise, excessive process disturbances and unrealistically fast controls have to be considered here.

No overshoot at start up: The most critical time for overshoot is the time of start up of a system. If we control an open tank, we do not want the tank to overflow as a result of overshoot of the level. More dramatically, if we have a closed tank, we do not want the tank to burst. Similar considerations exist everywhere, where danger of some sort exists.

Minimizing the effect of known disturbances: If we can measure disturbances, we may have a chance to control them before their effects become apparent.

3.4.1. Continuous cycling method (Ziegler Nichols)

This method of tuning requires determining the critical value of controller gain (K_C) that will produce a continuous oscillation of a control loop. This will occur when the total loop gain (K_{LOOP}) is equal to one. The controller gain value (K_C) then becomes known as the ultimate gain (K_U).

If we consider a basic liquid flow control loop utilizing:

- A venturi flow meter with a 4-20 mA output feeding into…
- a PID controller which in turn has a 4-20 mA output that controls...
- a valve actuator that in turn varies the flow rate of…
- the process.

When the product of the gains of all four of these component parts equals one, the system will become unstable when a process disturbance occurs (a set-point change). It will oscillate at its natural frequency which is determined by the process lag and response time, and caused by the loop gain becoming one.

Then measure the frequency of oscillation (the period of one cycle of oscillation), this being the ultimate period PU.

In addition, the final value of K_C is the critical gain of the controller (K_U). This gain value, when multiplied with the unknown process Gain(s), will give a Loop Gain, K_{LOOP}, of 1.

3.4.2. The stages of obtaining closed loop tuning (continuous cycling method)

- Put Controller in P-Control Only
- Select the P-Control to ERR = (SP - PV)
- Put the Controller into Automatic Mode
- Make a Step Change to the Set point
- Take action based on the Observation
- Conclude the Tuning Procedure.

3.4.3. Damped cycling tuning method

This method is a variation of the continuous cycling method. It is used whenever continuous cycling imposes a danger to the process, but a damped oscillation of some extent is acceptable.

The steps of closed loop tuning (damped cycling method) are as follows:

- Put the Controller into P-Control Only
- Select the P-Control to ERR = (SP - PV)
- Put the Controller in Automatic Mode
- Make a Step Change to the Set point
- Take action based on the Observation.

3.5. Cascade Control

If the OP of the temperature controller TC drives the SP of this newly added fuel flow controller FC, then there is a situation that the OP of the temperature controller TC then drives the true flow and not just a valve position.

Fuel flow pressure would practically have no effect on the outlet temperature. This concept is called 'cascade control'. The principle is shown in Figure 3.9.

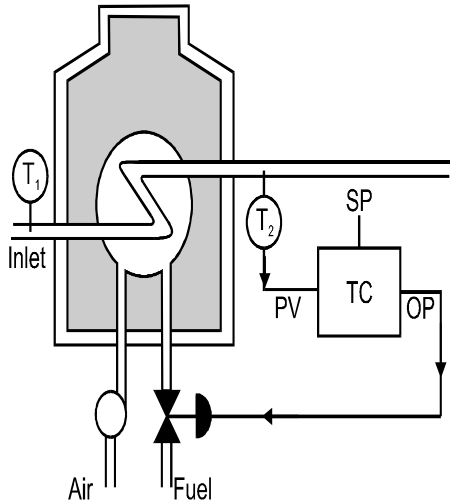


Figure 3.9
Single loop temperature control

3.5.1. The concept of process variable or PV-tracking

PV-Tracking is active if the secondary (FC) controller is in manual mode. Controllers can be set up to make use of PV-Tracking or not.

The concept is that an operator sets the OP value of the fuel controller manually until they find an appropriate value for the process.

3.6. Initialization of a cascade system

Initialization is actually a kind of manual mode where the operator does not drive the OP value of the primary controller. (temperature controller, TC, in this case.) Instead, fuel controller FC supplies its set point (SP) value, back up the cascade chain to the OP of the controller that will be driving it (the FC's SP) when the system is in automatic mode. If selected, PV-Tracking can take place in the primary controller as it would occur in normal manual mode.

3.7. Feed forward Control

If, within a process control's feedback system, large and random changes to either the PV or Lag time of the process occur, the feedback action becomes very ineffective in trying to correct these excessive variances.

These variances usually drive the process well outside its area of operation, and the feedback controller has little chance of making an accurate or rapid correction back to the SP term.

The result of this is that the accuracy and standard of the process becomes unacceptable. Feedforward control is used to detect and correct these disturbances before they have a chance to enter and upset the closed or feedback loop characteristics.

Feedforward Control has

- Manual feedforward control
- Automatic feedforward control

3.8. Manual feedforward control

Here, as a disturbance enters the process, it is detected and measured by the process operator. Based on his knowledge of the process, the operator then changes the manipulated variable by an amount that will minimize the effect of the measured disturbance on the system.

This form of feedforward control relies heavily on the operator and his knowledge of the operation of the process. However, if the operator makes a mistake or is unable to anticipate a disturbance, then the controlled variable will deviate from its desired value and, if feedforward is the only control, an uncorrected error will exist.

3.9. Automatic feedforward control

Disturbances that are about to enter a process are detected and measured. Feedforward controllers then change the value of their manipulated variables (outputs) based on these measurements as compared with their individual set-point values.

Feedforward controllers must be capable of making a whole range of calculations, from simpe on-off action to very sophisticated equations. These calculations have to take into account all the exact effects that the disturbances will have on controlled variables.

Pure feedforward control is rarely encountered; it is more common to find it embedded within a feedback loop where it assists the feedback controller function by minimizing the impact of excessive process disturbances.

3.10. Time matching as feedforward control

Time taken for a process to react in one direction (heating) is different to the time taken for the process to return to its original state (cooling). If the reaction curve (dynamic behavior of reaction) of the process disturbance is not equal to the control action, it has to be made equal.

Normally Lead/Lag compensators as tools are used to obtain equal dynamic behavior. They compensate for the different speeds of reaction. A problem of special importance is the drifting away of the PV. One can be as careful as one wants with evaluation of the disturbances, but never reach the situation of absolute perfect compensation. There are always factors not accounted for. This causes a drifting of the PV which has to be corrected manually from time to time, or an additional feedback control has to be added.

3.10.1. Process dead time

Overcoming the dead time in a feedback control loop can present one of the most difficult problems to the designer of a control system. This is especially true if the dead time is greater than 20% of the total time taken for the PV to settle to its new value after a change to the SP value of a system.

If the time from a change in the manipulated variable (controller output) and a detected change in the PV occurs, any attempt to manipulate the process variable before the dead time has elapsed will inevitably cause unstable operation of the control loop. Figure 3.10 illustrates various dead times and their relationship to the PV reaction time.

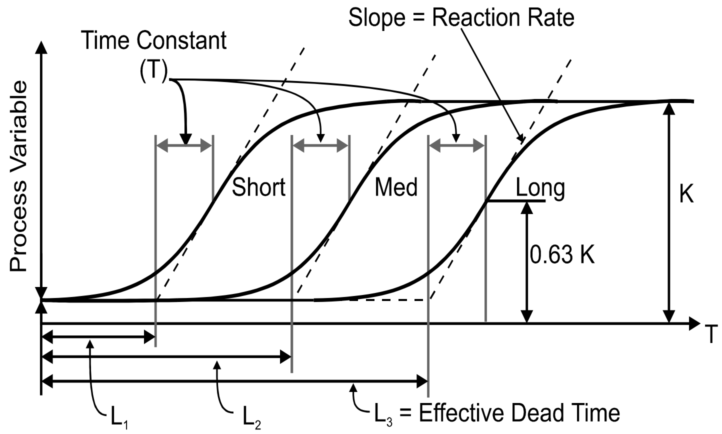


Figure 3.10
Reaction curves showing short, medium and long dead times

3.11. Overcoming Process dead time

Solving these problems depends to a great extent on the operating requirement(s) of the process. The easiest solution is to “de-tune” the controller to a slower response rate. The controller will then not overcompensate unless the dead time is excessively long.

The integrator (I mode) of the controller is very sensitive to “dead time” as during this period of inactivity of the PV (an ERR term is present) the integrator is busy “ramping” the output value.

Ziegler and Nichols determined the best way to “de-tune” a controller, to handle a dead time of D minutes, is to reduce the integral time constant T_{INT} by a factor of D^2 and the Proportional constant by a factor of D.

The derivative time constant T_{DER} is unaffected by dead time as it only occurs after the PV starts to move.

If, however, we could inform the controller of the dead time period, and give it the patience to wait and be content until the dead time has passed, then detuning and making the whole process very sluggish would not be required. This is what the smith predictor attempts to perform.

3.12. First term explanation(disturbance free PV)

The first term is an estimate of what the PV would be like in the absence of any process disturbances. It is produced by running the controller output through a model that is designed to accurately represent the behavior of the process without taking any load disturbances into account. This model consists of two elements connected in series.

- The first represents all of the process behavior not attributable to dead time. This is usually calculated as an ordinary differential or difference equation that includes estimates of all the process gains and time constants.
- The second represents nothing but the dead time and consists simply of a time delay, what goes in, comes out later, unchanged.

3.13. Second term explanation(predicted PV)

The second term introduced into the feedback path is an estimate of what the PV would look like in the absence of both disturbances and dead time. It is generated by running the controller output through the first element of the model (gains and TCs) but not through the time delay element.

It thus predicts what the disturbance-free PV will be like once the dead time has elapsed.

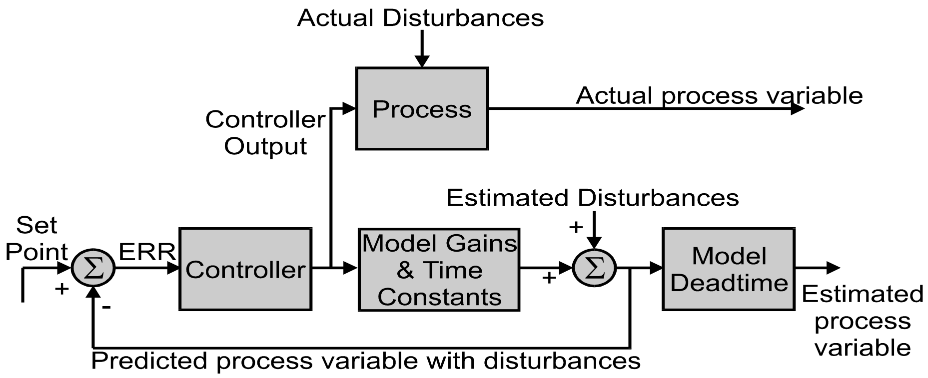


Figure 3.11
The smith predictor in use

If it is successful in doing so and the process model accurately emulates the process itself, then the controller will simultaneously drive the actual PV toward the SP value, irrespective of SP changes or load disturbances.

Chapter 4. Advanced Process Control

4.1. Introduction

Advanced process control (APC) is a broad term within the control theory. It is composed of different kinds of process control tools, for example, model predictive control (MPC), statistical process control (SPC), Run2Run (R2R), fault detection and classification (FDC), sensor control and feedback systems. APC is often used for solving multivariable control problems or discrete control problems.

4.2. Overview of Advanced Control Methods

4.2.1. Adaptive Control

An adaptive control system can be defined as a feedback control system intelligent enough to adjust its characteristics in a changing environment so as to operate in an optimal manner according to some specified criteria.

Generally speaking, adaptive control systems have achieved great success in aircraft, missile, and spacecraft control applications. It can be concluded that traditional adaptive control methods are mainly suitable for:

- Mechanical systems that do not have significant time delays; and
- Systems that have been designed so that their dynamics are well understood.

In industrial process control applications, however, traditional adaptive control has not been very successful.

4.2.2. Robust Control

Robust control is a controller design method that focuses on the reliability (robustness) of the control algorithm. Robustness is usually defined as the minimum requirement a control system has to satisfy to be useful in a practical

environment. Once the controller is designed, its parameters do not change and control performance is guaranteed.

Robust control methods are well suited to applications where the control system stability and reliability are the top priorities, process dynamics are known, and variation ranges for uncertainties can be estimated. Aircraft and spacecraft controls are some examples of these systems.

4.2.3. Predictive Control

Predictive control, or model predictive control (MPC), is one of only a few advanced control methods used successfully in industrial control applications. The essence of predictive control is based on three key elements:

- Predictive model,
- Optimization in range of a temporal window, and
- Feedback correction.

These three steps are usually carried on continuously by computer programs online. Predictive control is a control algorithm based on the predictive model of the process. The model is used to predict the future output based on the historical information of the process as well as the future input. It emphasizes the function of the model, not the structure of the model.

Predictive control is an algorithm of optimal control. It calculates future control actions based on a penalty function or performance function. The optimization of predictive control is limited to a moving time interval and is carried on continuously online. The moving time interval is sometimes called a temporal window. This is the key difference compared to traditional optimal control that uses a performance function to judge global optimization

Predictive control is also an algorithm of feedback control. If there is a mismatch between the model and process, or if there is a control performance problem caused by the system uncertainties, the predictive control could compensate for the error or adjust the model parameters based on on-line identification.

4.2.4. Optimal Control

Optimal control is an important component in modern control theory. Its great success in space, aerospace, and military applications has changed our lives in many ways.

The statement of a typical optimal control problem can be expressed in the following:

”The state equation and its initial condition of a system to be controlled are given. The defined objective set is also provided.”

Find a feasible control such that the system starting from the given initial condition transfers its state to the objective set, and minimizes a performance index. In practice, optimal control is very well suited for space, aerospace, and military applications such as the moon landing of a spacecraft, flight control of a rocket, and the missile blocking of a defense missile.

4.2.5. Intelligent Control

Intelligent control is another major field in modern control technology. There are different definitions regarding intelligent control, but it is referred to as a control Para diagram that uses various artificial intelligence techniques, which may include the following methods:

- Learning control,
- Expert control,
- Fuzzy control, and
- Neural network control.

Learning Control: Learning control uses pattern recognition techniques to obtain the current status of the control loop; and then makes control decisions based on the loop status as well as the knowledge or experience stored previously.

Expert Control: Expert control, based on the expert system technology, uses a knowledge base to make control decisions. The knowledge base is built by human expertise, system data acquired on-line, and inference machine designed. Since the knowledge in expert control is represented symbolically and is always in discrete format, it is suitable for solving decision making problems such as production planning, scheduling, and fault diagnosis. It is not well suited for continuous control issues.

Fuzzy Control: Fuzzy control, unlike learning control and expert control, is built on mathematical foundations with fuzzy set theory. It represents knowledge or experience in a mathematical format that process and system dynamic characteristics can be described by fuzzy sets and fuzzy relational functions. Control decisions can be generated based on the fuzzy sets and functions with rules.

Neural Network Control: Neural network control is a control method using artificial neural networks. It has great potential since artificial neural networks are built on a firm mathematical foundation that includes versatile and well understood mathematical tools. Artificial neural networks are also used as one of the key elements in the model-free adaptive controllers.

4.3. Internal Model Control

The Internal Model control (IMC) philosophy relies on the Internal Model principle, which states that "control can be achieved only if the control system encapsulates, either implicitly or explicitly; some representation of the process to be controlled". In particular, if the control scheme has been developed based on an exact model of the process, then perfect control is theoretically possible. Consider the example shown in the diagram below.

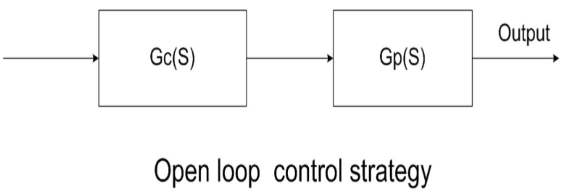


Figure 4.1
Open loop control strategy

A controller, $G_c(s)$, is used to control the process, $G_p(s)$. Suppose $\tilde{G}_p(s)$ is a model of $G_p(s)$. By setting $G_c(s)$ to be the inverse of the model of the process,

$$G_c(s) = \tilde{G}_p(s)^{-1},$$

And if $G_p(s) = \tilde{G}_p(s)$,(the model is an exact representation of the process)

Then it is clear that the output will always be equal to the set point.

4.3.1. The IMC Strategy

In practice, however, process-model mismatch is common; the process model may not be invertible and the system is often affected by unknown disturbances. Thus the above open loop control arrangement will not be able to maintain output at set point. Nevertheless, it forms the basis for the development of a control strategy that has the potential to achieve perfect control.

4.3.1.1. Model Predictive Control(MPC)

Model predictive control, or MPC, is an advanced method of process control. Model predictive controllers rely on dynamic models of the process, most often linear empirical models obtained by system identification. The models are used to predict the behavior of dependent variables (i.e, outputs) of a dynamical system with respect to changes in the process independent variables (i.e., inputs). In chemical processes, independent variables are most often set points of regulatory controllers that govern valve movement (e.g. valve positioners with or without flow, temperature or pressure controller cascades), while dependent variables are most often constraints in the process (e.g. product purity, equipment safe operating limits). The model predictive controller uses the models and current plant measurements to calculate future moves in the independent variables that will result in an operation that honors all independent and dependent variable constraints. The MPC then sends this set of independent variable moves to the corresponding regulatory controller set points to be implemented in the process.

4.3.1.2. Model Representations

MPC is widely adopted in the process industry as an effective means to deal with large multivariable constrained control problems. The main idea of MPC is to choose the control action by repeatedly solving online an optimal control problem. This aims at minimizing a performance criterion over a future horizon, possibly subject to constraints on the manipulated inputs and outputs, where the future behavior is computed according to a model of the plant.

Predictive Constrained Control: PID type controllers do not perform well when applied to systems with significant time-delay. Perhaps the best known technique for controlling systems with large time-delays is the Smith Predictor. It overcomes the debilitating problems of delayed feedback by using predicted future states of the output for control.

Multivariable Control: Most processes require the monitoring of more than one variable. Controller-loop interaction exists such that the action of one controller affects other loops in a multi-loop system. Depending upon the inter-relationship of the process variables, tuning each loop for maximum performance may result in system instability when operating in a closed-loop mode. Loops that have single

input single output (SISO) controllers may therefore not be suitable for these types of applications. These types of controllers are not designed to handle the effects of loop interactions.

A multivariable controller, whether it be a Multiple Input Single Output (MISO) or a Multiple Input Multiple Output (MIMO) is used for systems that have these types of interactions.

Model-Based Predictive Control: Model-Based Predictive Control technology utilizes a mathematical model representation of the process. The algorithm evaluates multiple process inputs, predicts the direction of the desired control variable, and manipulates the output to minimize the difference between target and actual variables. Strategies can be implemented in which multiple control variables can be manipulated and the dynamics of the models are changed in real time.

Dynamic Matrix Control: Dynamic Matrix Control (DMC) is also a popular model-based control algorithm. A process model is stored in a matrix of step or impulse response coefficients. This model is used in parallel with the on-line process in order to predict future output values based on the past inputs and current measurements.

Statistical Process Control: Statistical Process Control (SPC) provides the ability to determine if a process is stable over time, or, conversely, if it is likely that the process has been influenced by "special causes" which disrupt the process. Statistical Control Charts are used to provide an operational definition of a "special cause" for a given process, using process data.

SPC has been traditionally achieved by successive plotting and comparing a statistical measure of the variable with some user defined control limits. If the plotted statistic exceeds these limits, the process is considered to be out of statistical control. Corrective action is then applied in the form of identification, elimination or compensation for the assignable causes of variation. "On-line SPC" is the integration of automatic feedback control and SPC techniques. Statistical models are used not only to define control limits, but also to develop control laws that suggest the degree of manipulation to maintain the process under statistical control.

Chapter 5. Industrial Data Communications and Wireless

5.1. Introduction

Data communication involves the transfer of information from one point to another. Many communication systems handle analog data; examples are telephone systems, radio and television. Modern instrumentation is almost wholly concerned with the transfer of digital data.

Any communications system requires a transmitter to send information, a receiver to accept it, and a link between the two. Types of link include copper wire, optical fiber, radio and microwave.

Digital data is sometimes transferred using a system that is primarily designed for analog communication. A modem, for example, works by using a digital data stream to modulate an analog signal that is sent over a telephone line. Another modem demodulates the signal to reproduce the original digital data at the receiving end. The word 'modem' is derived from modulator and demodulator.

There must be mutual agreement on how data is to be encoded, i.e. the receiver must be able to understand what the transmitter is sending. The structure in which devices communicate is known as a protocol.

The standard that has created an enormous amount of interest in the past few years is Ethernet. Other protocol, which fits onto Ethernet extremely well, is TCP/IP, and being derived from the Internet is very popular and widely used.

5.2. Open Systems Interconnection (OSI) model

The OSI model, developed by the International Organization for Standardization, has gained widespread industry support. The OSI model reduces every design and communication problem into a number of layers as shown in Figure 5.1. A physical interface standard such as RS-232 would fit into the layer 1, while the other layers relate to the protocol software.

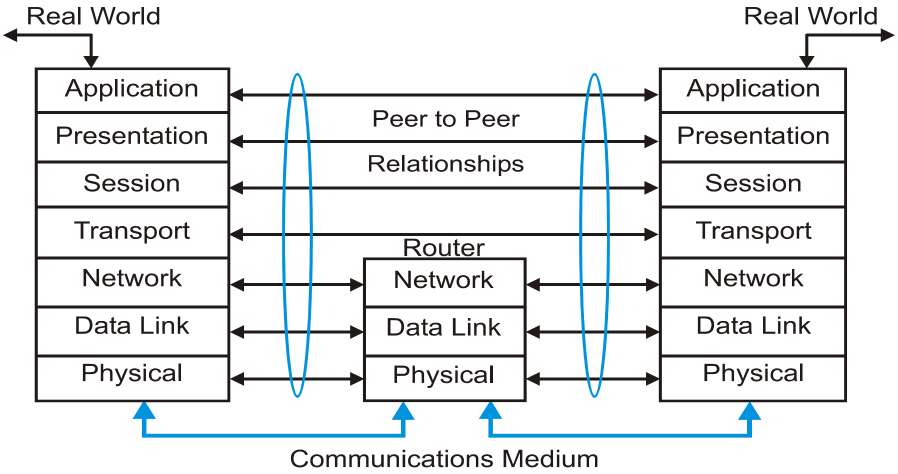


Figure 5.1
OSI model representation: two hosts interconnected via a router

The OSI model is useful in providing a universal framework for all communication systems. However, it does not define the actual protocol to be used at each layer. It is anticipated that groups of manufacturers in different areas of industry will collaborate to define software and hardware standards appropriate to their particular industry. Those seeking an overall framework for their specific communications' requirements have enthusiastically embraced this OSI model and used it as a basis for their industry specific standards.

5.2.1. Protocols

As previously mentioned, the OSI model provides a framework within which a specific protocol may be defined. A protocol, in turn, defines a frame format that might be made up of various fields as follows.

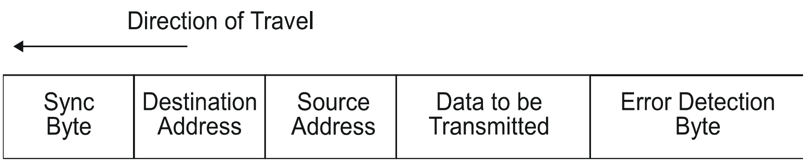


Figure 5.2
Basic structure of an information frame

5.3. RS-232 interface standard

The RS-232 interface standard (officially called TIA-232) defines the electrical and mechanical details of the interface between Data Terminal Equipment (DTE) and Data Communications Equipment (DCE), which employ serial binary data interchange. The current version of the standard refers to DCE as Data Circuit-terminating Equipment.

Figure 5.3 illustrates the signal flows across a simple serial data communications link.

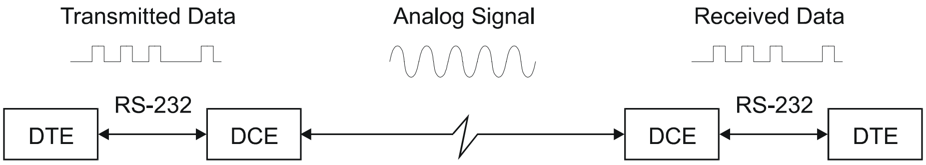


Figure 5.3
A typical serial data communications link

The RS-232 standard consists of three major parts, which define:

- Electrical signal characteristics
- Mechanical characteristics of the interface
- Functional description of the interchange circuits

5.3.1. Half-duplex operation of RS-232

The following description of one particular mode of operation of the RS-232 interface is based on half-duplex data interchange. The description encompasses the more generally used full-duplex operation.

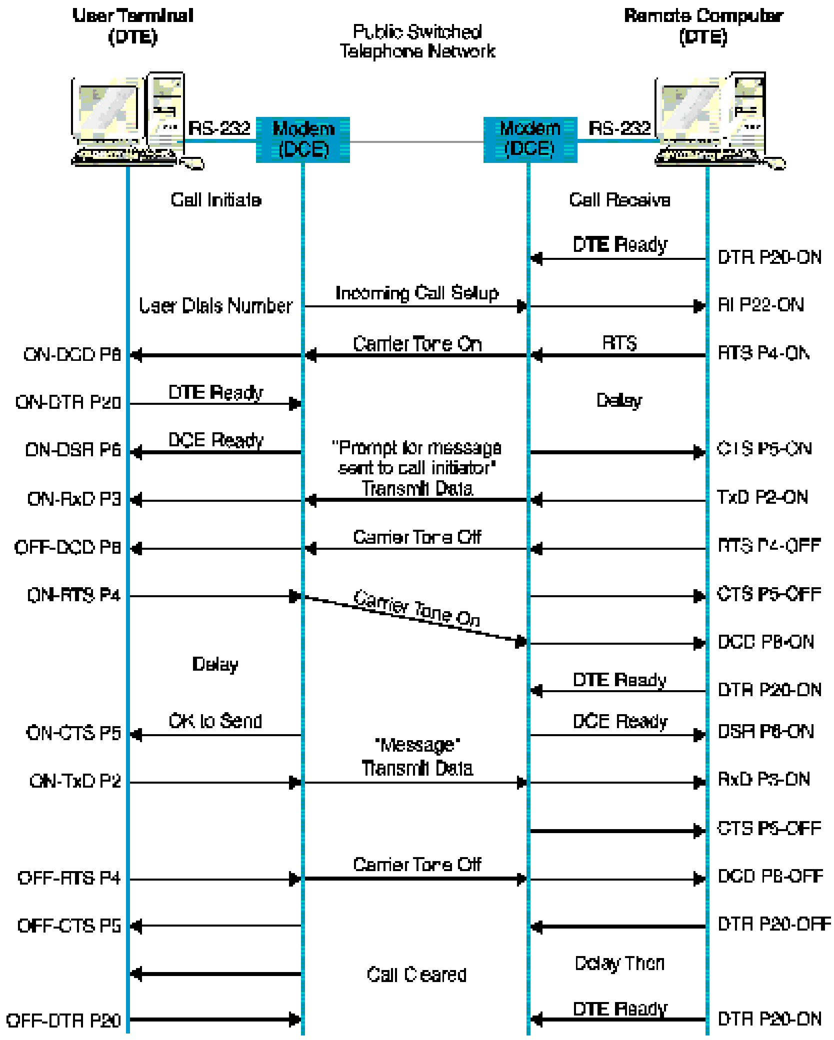


Figure 5.4
Half- duplex operational sequence of RS-232

Figure 5.4 shows the operation with the initiating user terminal, DTE, and its associated modem, DCE, on the left of the diagram and the remote computer and its modem on the right.

Full-duplex operation requires that transmission and reception must be able to occur simultaneously. In this case, there is no RTS/CTS interaction at either end. The RTS and CTS lines are left ON with a carrier to the remote computer.

5.4. Fiber Optics

Fiber optic communication uses light signals guided through a fiber core. Fiber optic cables act as waveguides for light, with all the energy guided through the central core of the cable. The light is guided due to the presence of a lower refractive index cladding around the central core. Little of the energy in the signal is able to escape into the cladding and no energy can enter the core from any external sources. Therefore the transmissions are not subject to any electromagnetic interference.

The core and the cladding will trap the light ray in the core, provided the light ray enters the core at an angle greater than the 'critical angle'. The light ray will then travel through the core of the fiber, with minimal loss in power, by a series of total internal reflections. Figure 5.5 illustrates this process.

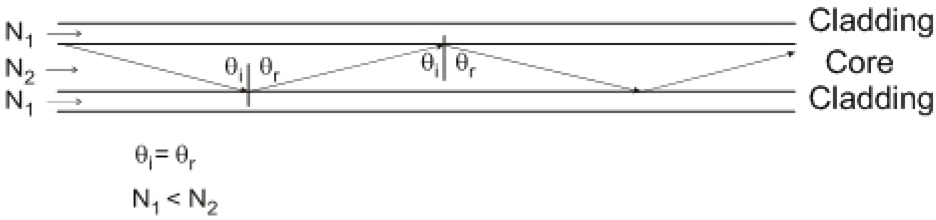


Figure 5.5
Light ray traveling through an optical fiber

5.4.1. Applications for fiber optic cables

Fiber optic cables offer the following advantages over other types of transmission media:

- Light signals are impervious to interference from EMI or electrical crosstalk
- Light signals do not interfere with other signals
- Optical fibers have a much wider, flatter bandwidth than coaxial cables and equalization of the signals is not required
- The fiber has a much lower attenuation, so signals can be transmitted much further than with coaxial or twisted pair cable before amplification is necessary
- Optical fiber cables do not conduct electricity and so eliminate problems of ground loops, lightning damage and electrical shock
- Fiber optic cables are generally much thinner and lighter than copper cables
- Fiber optic cables have greater data security than copper cables

5.4.2. Fiber optic cable components

The major components of a fiber optic cable are the core, cladding, coating (buffer), as shown in Figure 5.6. Some types of fiber optic cable even include a conductive copper wire that can be used to provide power to a repeater.

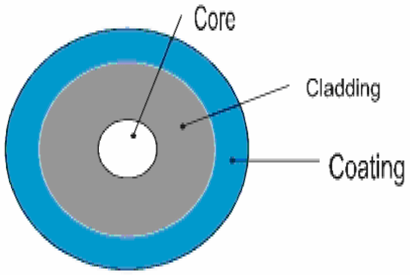


Figure 5.6
Fiber optic cable components

The fiber components include:

- Fiber core
- Cladding
- Coating (buffer)
- Strength members
- Cable sheath

There are four broad application areas into which fiber optic cables can be classified: aerial cable, underground cable, sub-aqueous cable and indoor cable.

5.5. Modbus

Modbus Messaging protocol is an Application layer (OSI layer 7) protocol that provides client/server communication between devices connected to different types of buses or networks. The Modbus Messaging protocol is only a protocol and does not imply any specific hardware implementation. Also note that the Modbus Messaging protocol used with Modbus Serial is the same one used with Modbus Plus and Modbus TCP.

Modbus messaging is based on a client/server model and employs the following messages:

- Modbus requests, i.e. the messages sent on the network by the clients to initiate transactions. These serve as indications of the requested services on the server side
- Modbus responses, i.e. the response messages sent by the servers. These serve as confirmations on the client side

The interaction between client and sever (controller and target device) can be depicted as follows. The parameters exchanged by the client and server consist of the Function Code ('what to do'), the Data Request ('with which input or output') and the Data response ('result').

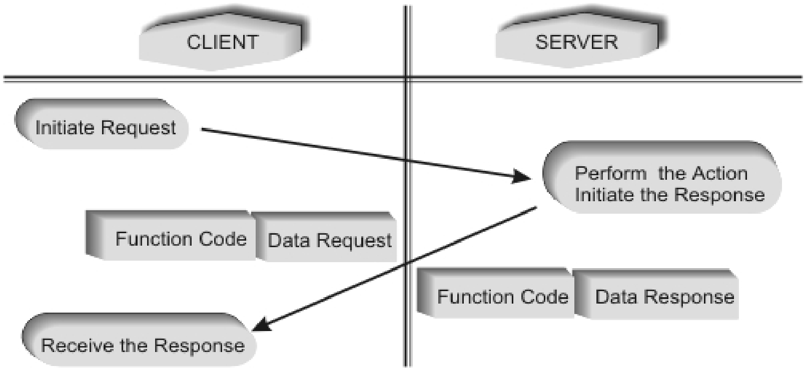


Figure 5.7
Modbus transaction

The Application Data Unit (ADU) structure of the Modbus protocol is shown in the Figure 5.8.

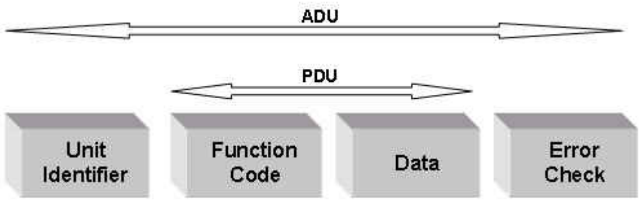


Figure 5.8
Modbus serial ADU format

Modbus functions can be divided into four groups or 'Conformance Classes'. The Function Codes are normally expressed in decimal; the hexadecimal equivalents are shown in brackets.

Conformance Class 0 is the minimum set of useful commands for both controllers and target devices. Note that the descriptions of certain commands have changed over the years, for this reason both the current and historical ('classic') descriptions are given here.

Table 5. 1
Conformance Class 0 commands

Function Code	Current terminology	Classic terminology
3 (0x03)	Read multiple registers	Read holding registers
16 (0x10)	Write multiple registers	Preset multiple registers

Conformance Class 1 comprises an additional set of commands, commonly implemented and interoperable.

Table 5. 2
Conformance Class 1 commands

Function Code	Current terminology	Classic terminology
1 (0x01)	Read coils	Read coil status
2 (0x02)	Read input discretes	Read input status
4 (0x04)	Read input registers	Read input registers

5(0x05)	Write coil	Write single register
6(0x06)	Force single coil	Preset single register
7(0x07)	Read exception status	Read exception status

Function Code 7 usually has a different meaning for each PLC family.

Conformance Class 2 comprises the data transfer functions needed for routine operations and supervision. These include, but are not limited to:

Table 5. 3
Conformance Class 2 commands

Function Code	Current terminology	Classic terminology
15 (0x0F)	Force multiple coils	Force multiple coils
22 (0x16)	Mask write register	Mask write register
23 (0x17)	Read/write registers	Read/write registers

There are also others such as Function Code 20 (read general reference), Function Code 21 (write general reference) and Function Code 24 (read FIFO queue) but they are considered to be outside the ambit of this section.

- Machine/vendor/network specific functions are those that, although being mentioned in the Modbus manuals, are not appropriate for interoperability because they are too machine-dependent. These include Function Codes such as 9 (program: Modicon 484), 10 (poll: Modicon 484) and 19 (reset communications link: Modicon 884/u84).

The following table summarizes the relationship of some of the more commonly used commands and the input/output addresses. The descriptions use the current rather than the classic terminology.

Table 5. 4
Modicums addresses and Function Codes

Data type	Absolute addresses	Relative addresses	Function codes	Description
Coils	00001 to 09999	0 to 9998	01	Read coils
Coils	00001 to 09999	0 to 9998	05	Write coil
Coils	00001 to 09999	0 to 9998	15	Write multiple coils
Discrete inputs	10001 to 19999	0 to 9998	02	Read input discretes
Input registers	30001 to 39999	0 to 9998	04	Read input registers
Holding registers	40001 to 49999	0 to 9998	03	Read multiple registers
Holding registers	40001 to 49999	0 to 9998	06	Write single register
Holding registers	40001 to 49999	0 to 9998	16	Write multiple registers
–	–	–	07	Read exception status
–	–	–	08	Loopback diagnostic test

5.5.1.1. Example of Function Code 2: Read input discretes

In classic terminology this function is known as 'read input status'. It enables the controller to read one or more discrete inputs in a target device. The data field of the request frame consists of the protocol address of the first discrete input followed by the number of discrete inputs to be read. The data field of the response frame consists of a count of the discrete input data bytes followed by that many bytes of discrete input data.

The discrete input data bytes are packed with one bit for the status of each consecutive discrete input. The least significant bit of the first discrete input data byte conveys the status of the first input read (i.e. the one with the lowest address). If the number of discrete inputs read is not an even multiple of eight, the last data byte will be padded with zeros on the high end. If there are more than eight bits in the response, the second byte will contain the next bits and so on. Once again this is not consistent with a big-endian approach.

In the following example, the controller requests the status of discrete inputs with protocol addresses 0x0000 and 0x0001 i.e. addresses 10001 and 10002 PLC. The target device's response indicates that discrete input 10001 is OFF and discrete input 10002 is ON (Figure 5.9).

- 'Reference number' refers to the input discrete with the lowest address
- 'Bit count' refers to the number of input discretes ('number of points') to be read and can vary between 1 and 2000
- 'Byte count' refers to the number of bytes required to return the requested input discrete values and is calculated as ((bit count + 7) / 8)

'Bit values' refer to the actual values of the individual inputs or 'input data'

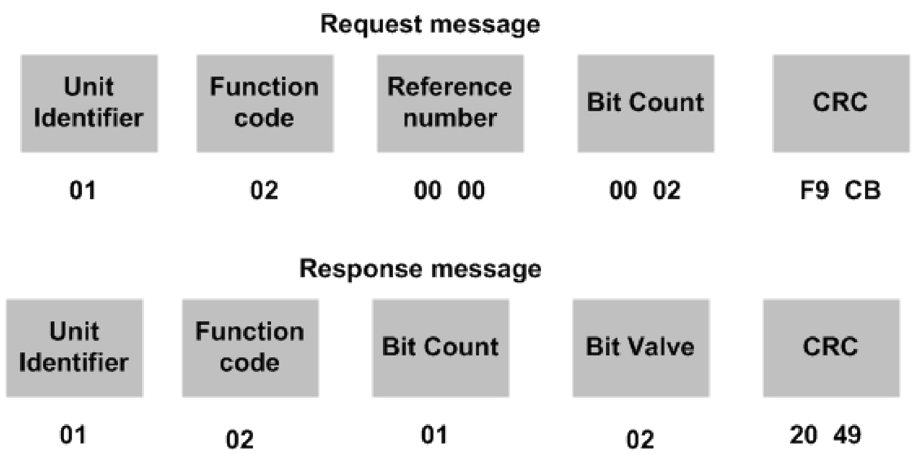


Figure 5.9
Example: FC02-reading input discretes

5.5.2. Modbus Plus

Modbus (or to be more exact; the Modbus Messaging protocol) is just a protocol, Modbus Plus is a complete system with a predefined medium and Physical layer (OSI layer 1) implementation. It is a LAN system for industrial control applications, allowing networked devices to exchange messages for the control

and monitoring of processes at remote locations in the industrial plant. Modbus Plus uses a token-passing medium access control mechanism, which results in deterministic operation, albeit not necessarily fast under all conditions.

The Modbus Plus layer 7 messaging protocol is essentially the same as that used for Modbus Serial and Modbus/TCP. The Physical layer is implemented with RS-485 and functions over shielded twisted pair cable. The Data Link layer (layer 2) protocol is based on the ISO/IEC 3309:1991 HDLC (High-level Data Link Control) multi-drop protocol, which uses a token passing medium access control mechanism and transmits data in a synchronous fashion as opposed to the asynchronous transmission of Modbus Serial. This results in transmission of data at 1 Mbps.

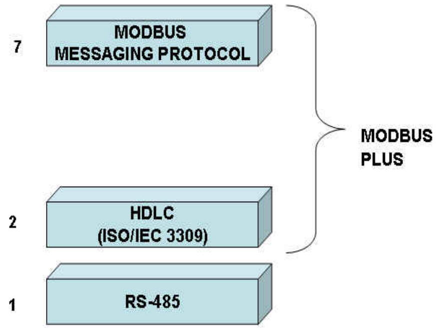


Figure 5.10
Modbus Plus protocol stack

Unlike Modbus, Modbus Plus is a proprietary standard developed to overcome the 'single-master' limitation prevalent in Modbus Serial.

5.6. Data Highway Plus /DH485

There are three main configurations used in Allen Bradley data communications:

Data Highway: This is a Local Area Network (LAN) that allows peer to peer communications amongst up to 64 nodes. It uses a half-duplex (polled) protocol and rotation of link mastership. It operates at 57.6kbaud.

Data Highway Plus: This is similar to the Data Highway network although is designed for fewer nodes, and operates at a data rate of 57.6kbaud. It has peer to peer communications with a token passing scheme to rotate link mastership among the nodes.

Note that both systems implement peer to peer communications through a modified token passing system called the 'floating master'. This is a fairly efficient mechanism as each node has an opportunity to become a master, at which time it can immediately transmit without risking contention on the bus. Both systems use a differential signaling system similar to RS-485.

The Allen Bradley Data Highway plus implements three layers of the OSI layer model:

- Physical layer hardware
- Data Link layer protocol

- Application layer protocol

Data Highway-485: This is used by the SLC range of Allen Bradley controllers and is based on RS-485.

5.7. HART

The HART system (and its associated protocol) was originally developed by Rosemount and is regarded as an open standard, available to all manufacturers. Its main advantage is that it enables the retention of the existing 4-20mA instrumentation cabling whilst using, simultaneously, the same wires to carry digital information superimposed on the analog signal.

HART is a hybrid analog and digital system, as opposed to most field bus systems, that are purely digital. It uses a Frequency Shift Keying (FSK) technique based on the Bell 202 standard. Two individual frequencies of 1200 and 2200 Hz, representing digits '1' and '0' respectively, are used. The average value of the 1200/2400Hz sine wave superimposed on the 4-20mA signal is zero; hence, the 4-20mA analog information is not affected.

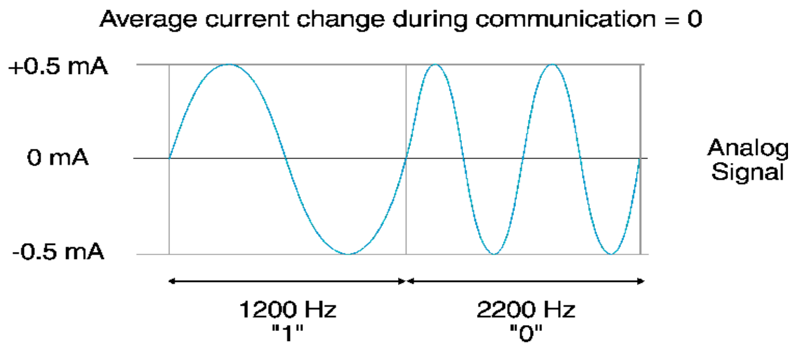


Figure 5.11
Frequency allocation of HART signaling system

HART can be used in three ways:

- In conjunction with the 4-20mA current signal in point-to-point mode
- In conjunction with other field devices in multi-drop mode
- In point-to-point mode with only one field device broadcasting in burst mode

Traditional point-to-point loops use zero for the smart device polling address. Setting the smart device polling address to a number greater than zero implies a multi-drop loop. Obviously the 4-20mA concept only applies to a loop with a single transducer; hence for a multi-drop configuration the smart device sets its analog output to a constant 4mA and communicates only digitally.

The HART protocol has two formats for digital transmission of data:

- Poll/response mode
- Burst (broadcast) mode

In the poll/response mode, the master polls each of the smart devices on the highway and requests the relevant information. In burst mode the field device

continuously transmits process data without the need for the master to send request messages. Although this mode is fairly fast (up to 3.7 times/second), it cannot be used in multidrop networks. The protocol is implemented with the OSI model using layers 1, 2 and 7.

5.8. AS-i

Actuator Sensor-interface is an open system network developed by eleven manufacturers.

AS-i is a bit-oriented communication link designed to connect binary sensors and actuators. Most of these devices do not require multiple bytes to adequately convey the necessary information about the device status, so the AS-i communication interface is designed for bit-oriented messages in order to increase message efficiency for these types of devices. It was not developed to connect intelligent controllers together since this would be far beyond the limited capability of such small message streams.

Modular components form the central design of AS-i. Connection to the network is made with unique connecting modules that require minimal, or in some cases no tools to provide for rapid, positive device attachment to the AS-i flat cable. Provision is made in the communications system to make 'live' connections, permitting the removal or addition of nodes with minimum network interruption.

Connection to higher level networks (e.g. ProfiBus) is made possible through plug-in PC and PLC cards or serial interface converter modules.

5.9. DeviceNet

DeviceNet, developed by Allen Bradley, is a low-level device oriented network based on CAN (Controller Area Network) developed by Bosch (GmbH) for the automobile industry. It is designed to interconnect lower level devices (sensors and actuators) with higher level devices (controllers). DeviceNet is classified as a field bus, per specification IEC-62026.

The variable, multi-byte format of the CAN message frame is well suited to this task as more information can be communicated per message than with bit-type systems. The DeviceNet specification is an open specification and available through the ODVA.

DeviceNet can support up to 64 nodes, which can be removed individually under power and without severing the trunk line. A single, four-conductor cable (round or flat) provides both power and data communications. It supports a bus (trunk line drop line) topology, with branching allowed on the drops. Reverse wiring protection is built into all nodes, protecting them against damage in the case of inadvertent wiring errors. The data rates supported are 125, 250 and 500K baud (i.e. bits per second in this case).

Figure 5.12 illustrates the positioning of DeviceNet and CANBUS within the OSI model. CANBUS represents the bottom two layers in the lower middle column, just below DeviceNet Transport. Unlike most other field buses, DeviceNet does implement layers 3 and 4, which makes it a routable system. There are two other products in the same family; Control Net and Ethernet/IP. They share the same

upper layer protocols (implemented by CIP, the Control and Information Protocol) and only differ in the lower four layers.

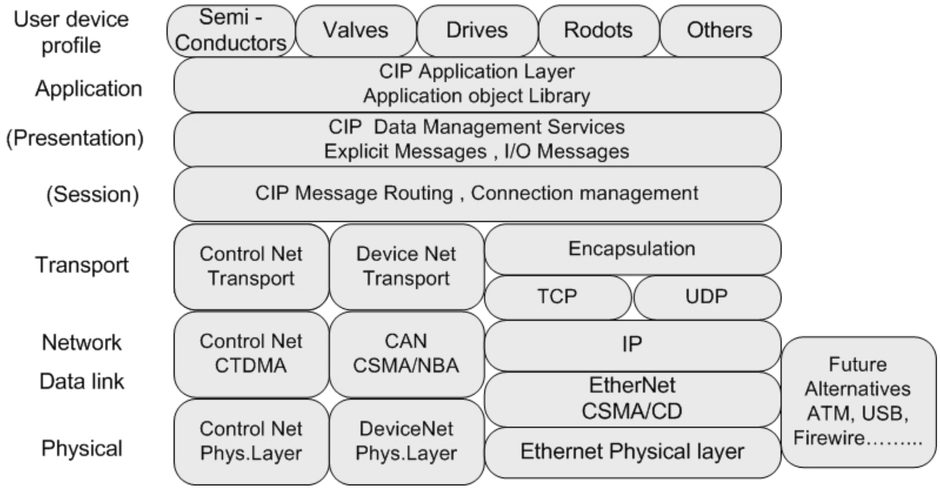


Figure 5.12
Devicenet (as well as ControlNet and Ethernet/IP) vs. the OSI model

5.10. Profibus

ProfiBus (**PRO**cess **FI**eld **BUS**) is a widely accepted international networking standard, commonly found in process control and in large assembly and material handling machines. It supports single-cable wiring of multi-input sensor blocks, pneumatic valves, complex intelligent devices, smaller sub-networks (such as AS-i), and operator interfaces.

It is an open, vendor independent standard. It adheres to the OSI model, ensuring that devices from a variety of different vendors can communicate easily and effectively. It has been standardized under the German National standard as DIN 19 245 Parts 1 and 2 and, in addition, has also been ratified under the European national standard EN 50170 Volume 2.

The bus interfacing hardware is implemented on ASIC (Application Specific Integrated Circuit) chips produced by multiple vendors, and are based on RS-485 as well as the European EN50170 Electrical specification.

ProfiBus uses 9-Pin D-type connectors (impedance terminated) or 12mm round (M12-style) quick-disconnect connectors. The number of nodes is limited to 127. The distance supported is up to 24km (with repeaters and fiber optic transmission), with speeds varying from 9600bps to 12Mbps. The message size can be up to 244 bytes of data per node per message (12 bytes of overhead for a maximum message length of 256 bytes), while the medium access control mechanisms are polling and token passing. ProfiBus supports two main types of devices, namely, masters and slaves.

- Master devices control the bus and when they have the right to access the bus, they may transfer messages without any remote request. These are referred to as active stations
- Slave devices are typically peripheral devices i.e. transmitters/sensors and actuators. They may only acknowledge

received messages or, at the request of a master, transmit messages to that master. These are also referred to as passive stations.

5.11. Foundation Fieldbus

Foundation Fieldbus allows end-user benefits such as:

- Reduced wiring
- Communications of multiple process variables from a single instrument
- Advanced diagnostics
- Interoperability between devices of different manufacturers
- Enhanced field level control
- Reduced start-up time
- Simpler integration.

The concept behind Foundation Fieldbus is to preserve the desirable features of the present 4-20mA standard while taking advantage of the new digital technologies. This provides the features noted above because of:

- Reduced wiring due to the multi-drop capability
- Flexibility of supplier choices due to interoperability
- Reduced control room equipment due to distribution of control functions to the device level
- Increased data integrity and reliability due to the application of digital communications.

Foundation Fieldbus implements four OSI layers. Three of them correspond to OSI layers 1, 2 and 7. The fourth is the so-called ‘user layer’ that sits on top of layer 7 and is often said to represent OSI ‘layer 8’.The user layer provides a standardized interface between the application software and the actual field devices.

5.12. Industrial Ethernet

Early Ethernet systems (of the 10 Mbps variety) use the CSMA/CD access method. This gives a system that operates with little delay if lightly loaded, but becomes very slow if heavily loaded. Ethernet network interface cards are relatively cheap and produced in vast quantities. Ethernet has, in fact, become the most widely used networking standard. However, CSMA/CD is a probabilistic medium access mechanism, there is no guarantee of message transfer and messages cannot be prioritized.

Modern Ethernet systems are a far cry from the original design. From 100BaseT onwards they are capable of full duplex (sending and receiving at the same time via switches, without collisions) and the Ethernet frame can be modified to make provision for prioritization and virtual LANs.

Early Ethernet was not entirely suitable for control functions as it was primarily developed for office-type environments. Ethernet technology has, however, made rapid advances over the past few years. It has gained such widespread acceptance in Industry that it is becoming the de facto field bus technology for OSI layers 1 and 2. An indication of this trend is the inclusion of Ethernet as the level 1 and 2

infrastructure for Modbus/TCP (Schneider), Ethernet/IP (Rockwell Automation and ODVA), ProfiNet (Profibus) and Foundation Fieldbus HSE.

5.12.1. 10 Mbps Ethernet

The IEEE 802.3 standard (also known as ISO 8802.3) defines a range of media types that can be used for a network based on this standard such as coaxial cable, twisted pair cable and fiber optic cable. It supports various cable media and transmission rates at 10 Mbps, such as:

- 10Base2 : thin wire coaxial cable (RG-58), 10 Mbps baseband operation, bus topology
- 10Base5 : thick wire coaxial cable (RG-8), 10 Mbps baseband operation, bus topology
- 10BaseT : UTP cable (Cat3), 10 Mbps baseband operation, star topology
- 10BaseFL : optical fiber, 10 Mbps baseband operation, point-to-point topology

Other variations included 1Base5, 10BaseFB, 10BaseFP and 10Broad36, but these versions never became commercially viable.

5.12.2. 100 Mbps Ethernet

100BaseT is the shorthand identifier for 100 Mbps Ethernet systems, viz. 100BaseTX (copper) and 100BaseFX (fiber). 100BaseT4 was designed to operate at 100 Mbps over 4 pairs of Cat3 cable, but this option never gained widespread acceptance. Yet another version, 100BaseT2, was supposed to operate over just 2 pairs of Cat3 cable but was never implemented by any vendor.

One of the limitations of hub-based (CSMA/CD) 100BaseT systems is the size of the collision domain, which is only 250 meters or 5.12 microseconds. This is the maximum size of a network segment in which collisions can be detected, being one tenth of the maximum size of a 10 Mbps network. This effectively limits the distance between a workstation and hub to 100 m, the same as for 10BaseT. As a result, networks larger than 200 meters must be logically interconnected by store-and-forward devices such as bridges, routers or switches. This is not a bad thing, since it segregates the traffic within each collision domain, reducing the number of collisions on the network. The use of bridges and routers for traffic segregation, in this manner, is often done on industrial Ethernet networks. Of course, the use of switches instead of hubs allows the construction of very large networks because of the full duplex operation.

The format of the frame has been left unchanged. The only difference is that it is transmitted 10 times faster than in 10 Mbps Ethernet, hence its length (in time) is 10 times less.

5.12.3. Gigabit Ethernet

1000BaseX is the shorthand identifier for the Gigabit Ethernet system based on the 8B/10B block encoding scheme adapted from the fiber channel networking standard, developed by ANSI. 1000BaseX includes 1000BaseSX, 1000BaseLX and 1000BaseCX.

- 1000BaseSX is the short wavelength fiber version
- 1000BaseLX is the long wavelength fiber version
- 1000BaseCX is a short copper cable version, based on the fiber channel standard.

1000BaseT, on the other hand, is a 1000 Mbps version capable of operating over Cat5 (or better, such as Cat5e) UTP, and has largely replaced 1000BaseCX. 1000BaseT is based on a different encoding scheme.

As with Fast Ethernet, Gigabit Ethernet supports full duplex and auto-negotiation. It uses the same frame format as 10 Mbps and 100 Mbps Ethernet systems, and operates at ten times the clock speed of Fast Ethernet, i.e. at 1Gbps. By retaining the same frame format as the earlier versions of Ethernet, backward compatibility is assured.

Despite the similar frame format, the system had to undergo a small change to enable it to function effectively at 1Gbps in CSMA/CD mode. The slot time of 64 bytes used with both 10 Mbps and 100 Mbps systems had to be increased by a factor of 8, to 512 bytes. This is equivalent to 4.096 μs. Without this increased slot time the collision domain would have been impracticably small at 25 meters. The irony is that in practice all Gigabit Ethernet systems are full duplex, and do not need this large slot time.

5.13. TCP/IP

TCP/IP is the de facto global standard for the Internet (network) and host–to–host (transport) layer implementation of internet work applications because of the popularity of the Internet. The Internet (known as ARPANet in its early years), was part of a military project commissioned by the Advanced Research Projects Agency (ARPA), later known as the Defense Advanced Research Agency or DARPA. The communications model used to construct the system is known as the ARPA model.

Whereas the OSI model was developed in Europe by the International Standards Organization (ISO), the ARPA model (also known as the DoD model) was developed in the USA by ARPA. Although they were developed by different bodies and at different points in time, both serve as models for a communications infrastructure and hence provide ‘abstractions’ of the same reality. The remarkable degree of similarity is therefore not surprising.

Whereas the OSI model has 7 layers, the ARPA model has 4 layers. The OSI layers map onto the ARPA model as follows.

- The OSI session, presentation and applications layers are contained in the ARPA process and application layer.
- The OSI transport layer maps onto the ARPA host–to–host layer (sometimes referred to as the service layer).
- The OSI network layer maps onto the ARPA Internet layer.
- The OSI physical and data link layers map onto the ARPA network interface layer.

The relationship between the two models is depicted in Figure 5.13.

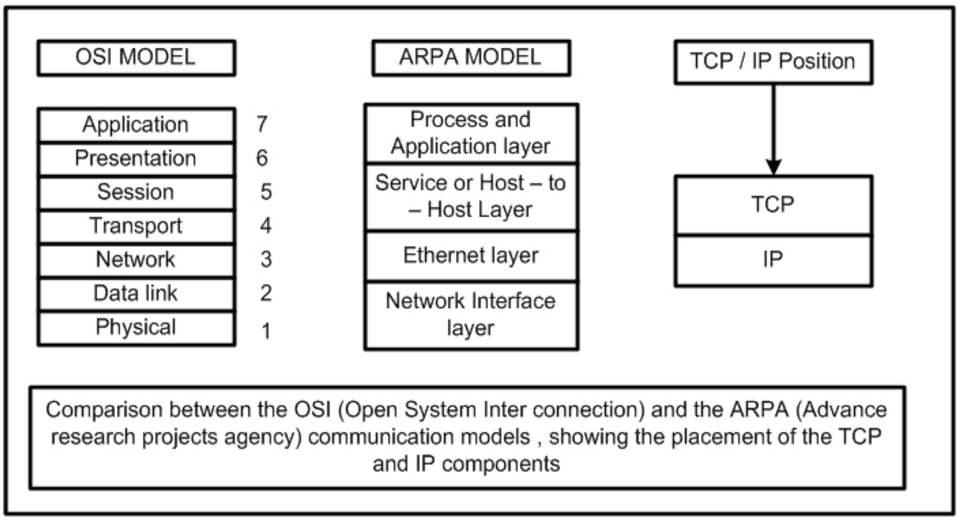


Figure 5.13
OSI vs ARPA models

TCP/IP, or rather the TCP/IP protocol suite is not limited to the TCP and IP protocols, but consists of a multitude of interrelated protocols that occupy the upper three layers of the ARPA model. TCP/IP does NOT include the bottom network interface layer, but depends on it for access to the medium.

As depicted in Figure 5.14, an Internet transmission frame originating on a specific host (computer) would contain the local network (for example, Ethernet) header and trailer applicable to that host. As the message proceeds along the Internet, this header and trailer could be replaced depending on the type of network on which the packet finds itself - be that X.25, frame relay or ATM. The IP datagram itself would remain untouched, unless it has to be fragmented and reassembled along the way.

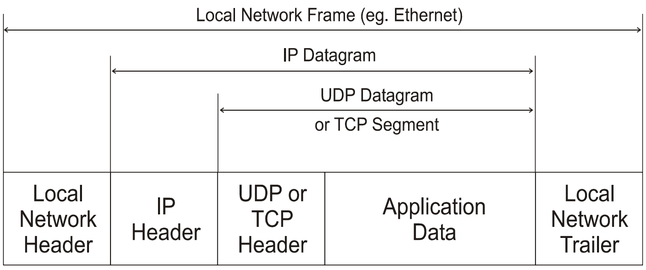


Figure 5.14
Internet frame

The Internet layer: This layer is primarily responsible for the routing of packets from one host to another.

The host–to–host layer: This layer is primarily responsible for data integrity between the sender host and receiver host regardless of the path or distance used to convey the message.

The process/application layer: This layer provides the user or application programs with interfaces to the TCP/IP stack.

Internet layer protocols (packet transport): Protocols like internet protocol (IP), the internet control message protocol (ICMP) and the address resolution protocol (ARP) are responsible for the delivery of packets (datagrams) between hosts.

Routing: Unlike the host–to–host layer protocols (for example, TCP), which control end–to–end communications, IP is rather 'shortsighted.' Any given IP node (host or router) is only concerned with routing (switching) the datagram to the next node, where the process is repeated.

5.14. Wireless Fundamentals

Wireless communication is the transfer of information over a distance without the use of electrical conductors or "wires”. The distances involved may be short (a few meters as in a television remote control) or very long (thousands or even millions of kilometers for radio communications). The term wireless technology is generally used for mobile IT equipment. It encompasses cellular telephones, personal digital assistants (PDAs), and wireless networking. Other examples of wireless technology include GPS units, garage door openers and/or garage doors, wireless computer mice and keyboards, satellite television and cordless telephones

Wireless communication involves:

- Radio frequency communication,
- Microwave communication, for example long-range line-of-sight via highly directional antennas, or short-range communication, or
- Infrared (IR) short-range communication, for example from remote controls or via IRDA.

Applications may involve point-to-point communication, point-to-multipoint communication, broadcasting , cellular networks and other wireless networks.

In the last 50 years, the wireless communications industry experienced drastic changes driven by many technology innovations. And quite often, there are start-up companies emerging and growing into multi-nationals.

Examples of wireless technology at work :

- Security systems
- Television remote control
- Cellular telephones.

Wireless is a term used to describe telecommunications in which electromagnetic waves (rather than some form of wire) carry the signal over part or the entire communication path. Common examples of wireless equipment in use today include:

- Cellular phones and pagers
- Global Positioning System (GPS)
- Cordless computer peripherals
- Cordless telephone sets
- Satellite television.

Wireless networking is used to meet a variety of needs. Perhaps the most common use is to connect laptop users who travel from location to location. Another common use is for mobile networks that connect via satellite. A wireless

transmission method is a logical choice to network a LAN segment that must frequently change locations. The following situations justify the use of wireless technology:

- To span a distance beyond the capabilities of typical cabling
- To avoid obstacles such as physical structures, EMI, or RFI
- To provide a backup communications link in case of normal network failure
- To link portable or temporary workstations
- To overcome situations where normal cabling is difficult or financially impractical, or
- To remotely connect mobile users or networks

5.15. Radio/microwave communications

A significant number of industrial protocols are transferred using radio telemetry systems. Radio is often selected in preference to using landlines for a number of reasons:

- Costs of cables and laying can far exceed that of radio telemetry systems
- Radio systems can be installed faster than landline systems
- Radio equipment is very portable and can be easily moved
- Radio can be used to transmit the data in any format required by the user
- Reasonably high data rates can be achieved compared to some landline applications
- Radio can be used as back up for landlines

The various aspects of radio and microwave communications that demand further detail in discussion are listed below:

- Components of a radio link
- Radio spectrum and frequency allocation
- Summary of radio characteristics for VHF/UHF radio telemetry systems
- Radio modems
- How to prevent inter-modulation problems
- Implementing a radio link
- Miscellaneous considerations

5.16. Installation and Troubleshooting

When troubleshooting a communications system, the engineer or the technician tries to use some standard format to arrive at a quicker solution. Industrial communications systems do not always respond to the tried and tested approaches that worked with hardwired inputs and outputs.

Common problems and solutions: Some of the causes for industrial communications problems include:

- No power to the station on the network, resulting in a breakdown in communications
- Cable damage, with a resultant interruption in communications

- Earthing and grounding problems resulting in intermittent failure of communications
- Electrostatic damage to the communications ports
- Software crash on one of the stations resulting in communications failure
- High levels of electrostatic/electromagnetic interference on the communications link
- High traffic loads on the link, resulting in intermittent communications
- Electrical surge or transient through the communications system resulting in hardware damage

The impact on the communications system ranges from outright failure (with no communications possible) to intermittent communications depending on the severity of the problem. Intermittent failure is arguably the worst problem to have, as it is very difficult to diagnose and fix.

General comments on troubleshooting: Obviously, there is no cut and dried method of testing. It depends on the environment and the history of the system. However, a few rules are useful in troubleshooting a communications system effectively.

- Extensive and accurate documentation
- Baseline reporting
- Network simplification

A specific methodology: When troubleshooting your communications system, the following steps should be taken:

- Check that all stations and network communications devices are powered up and operational
- Check all cabling for clean connections.
- Check grounding and earthing setups.
- Some new devices operating on the same power supply may be the cause of the problem
- Check whether there has been any changes or damage to screening of the cables.
- Use the diagnostics packages provided as part of the system to compare the number of packets transmitted to packets dropped.
- Commence by removing devices that are not critical to the system under investigation.
- Do simple diagnostic tests using simple utilities such as ‘ping’ or ‘netstat’ to identify what is happening on the network

5.16.1. RS-232

Since RS-232 is a point-to-point system, installation is fairly straightforward and all RS-232 devices use either DB-9 or DB-25 connectors. These connectors are cheap and allow multiple insertions. None of the RS-232 standards define which device uses a male or female connector, but traditionally the male (pin) connector is used on the DTE and the female connector (socket) is used on DCE equipment. This is only traditional and may vary on different equipment. It is often asked why a 25-pin connector is used when only 9 pins are needed. This was done because

RS-232 was used before the advent of computers and therefore used for hardware control (RTS/CTS). It was originally thought that, in the future, more hardware control lines would be needed hence the need for more pins.

During an installation of RS-232 connection, it is important to ask the following questions:

- Is one device a DTE and the other a DCE?
- What is the gender and size of connectors at each end?
- What is the speed of the communication?
- What is the distance between the equipment?
- Is it a noisy environment?
- Is the software set up correctly (all the UART parameters the same for both sides)?

5.16.2. RS-485

The RS-485 line drivers/receivers are differential chips. This means that the A and B wires are referenced to each other. A 'one' is transmitted, for example, when one of the lines is at +5V and the other one is at 0V. A 'zero' is then transmitted when the line voltages are reversed. In working systems the voltages are usually somewhere around +/- 2V with reference to each other. Up to 32 devices can be connected on one system without a repeater. Some systems allow the connection of five legs with four repeaters and get 160 devices on one system.

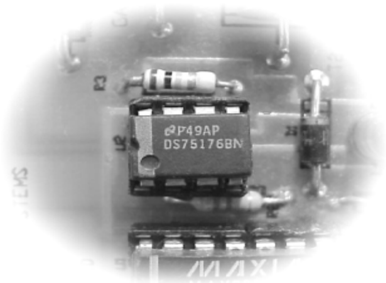


Figure 5.15
RS-485 chip

Note: ProfiBus DP and FMS use RS-485 at the Physical layer and therefore all the RS-485 installation and troubleshooting guidelines apply.

5.16.3. Modbus

No matter what extreme care you may have taken, there is hardly ever an installation that experiences trouble-free setup and configuration. Some common problems related to Modbus installations are listed below. They can be categorized as either hardware or software problems.

- Hardware problems include mis-wired communication cabling and faulty communication interfaces
- Software (protocol) related issues arise when the controller application tries to access non-existent target devices' nodes or uses invalid Function Codes, addresses non-existent memory locations in

the target devices, or specifies illegal data format types, which obviously the target devices do not understand.

5.16.4. Modbus plus

The Modbus Plus network is a 3-wire (one pair and a shield) twisted pair cable with the nodes connected in a daisy-changed configuration. There is no polarity requirement at the node's transceiver, so the data cable pair may be connected either way at a node. A 220-ohm terminator is required at each end of the network cable. There are limits on the maximum number of nodes per segment, the number of repeaters, and the lengths of cable segments on the Modbus Plus network.

The node address of the Modbus Plus device should be set before connecting it to the network. This avoids possible duplicate address problems with other units on the network.

Most software related issues arise from the use of invalid target device addressing; illegal target memory addressing, illegal data formats and even perhaps use of unrecognized function codes. Other issues are related to the actual configuration of the communication hardware itself.

5.16.5. Data Highway

Data Highway Plus wiring troubleshooting: Inspect the cable closely for wiring problems if the operation of the network appears intermittent. Typical problems include:

- Damage to the cable
- No terminator (150 Ω) at the end of the line
- Screen that are not grounded or damaged.

Data Highway Plus network diagnostics: Many of the errors are the result of excessive noise on the network and can be corrected by examining the actual wiring and removing the source of noise, if possible. If not (for example, due to the highway (trunk line) parallel to a power cable in a cable tray), consideration will have to be given to the use of fiber cabling as a replacement for the copper cable.

A few errors (identified in the diagnostics registers on the interface module) worth mentioning are:

- ACK Time out
- Contention
- False poll
- Transmitted messages and received messages
- Data pin allocation.

Note that the rules for troubleshooting the physical side of these two cables are very similar to that for RS-485. In fact, DH485 is identical to RS-485 while Data Highway Plus is essentially a transformer isolated version.

The difficult part in diagnosing problems with Data Highway Plus is in the operation of the protocol.

5.16.6. HART

Beside the actual instruments that require calibration, the only major problem that can occur with HART is the cable length calculation. The HART protocol is designed to work over existing analog signal cables, but it depends on sufficient voltage drop across the series resistor. This, in turn, depends on:

- The series load resistor
- Cable resistance
- Cable capacitance
- The number and total capacitance of the field devices
- The resistance of, and position of other devices in the loop

The main reason for this is that network must pass the HART signal frequencies without excessive loss or distortion. A software package such as H-Sim can be used to calculate whether the system is operating with the correct signal level. In addition, it should be confirmed that the loop has a bandwidth of at least 2500 Hz. This can be achieved by checking that the product of the cable resistance and capacitance (R times C) is less than 65 microseconds.

5.16.7. AS-i

The AS-i system has been designed with a high degree of 'maintenance friendliness' in mind and has a high level of built-in auto-diagnosis. The system is continuously monitoring itself against faults such as:

- Operational slave errors (permanent or intermittent slave failure, faulty configuration data such as addresses, I/O configuration, and ID codes)
- Operational master errors (permanent or intermittent master failure, faulty configuration data such as addresses, I/O configuration, and ID codes)
- Operational cable errors (short circuits, cable breakage, corrupted telegrams due to electrical interference and voltage outside of the permissible range)
- Maintenance related slave errors (false addresses entered, false I/O configuration, false ID codes)
- Maintenance related master errors (faulty projected data such as I/O configuration, ID codes, parameters etc.)
- Maintenance related cable errors (counter poling the AS-i cable)

The fault diagnosis is displayed by means of LEDs on the master. Where possible, the system will protect itself during short-circuit.

5.16.8. DeviceNet

Networks, in general, exhibit the following types of problems from time to time. The first type of problem is of an electronic nature, where a specific node (e.g. a network interface card) malfunctions. This can be due to a component failure or to an incorrect configuration of the device.

The second type is related to the medium that interconnects the nodes. Here, the problems are more often of an electromechanical nature and include open and short circuits, electrical noise, signal distortion and attenuation. Open and short

circuits in the signal path are caused by faulty connectors or cables. Electrical interference (noise) is caused by incorrect grounding, broken shields or external sources of electro-magnetic or radio frequency interference. Signal distortion and attenuation can be caused by incorrect termination, failure to adhere to topology guidelines (e.g. drop cables too long), or faulty connectors.

Whereas these are general network-related problems, the following ones are very specific to Devicenet:

- Missing terminators
- Excessive common mode voltage, caused by faulty connectors or excessive cable length
- Low power supply voltage caused by faulty connectors or excessive cable length
- Excessive signal propagation delays caused by excessive cable length

5.16.9. Ethernet

Ethernet hardware is fairly simple and robust, and once a network is commissioned with professional cabling and certification, the network should be fairly trouble-free. Most problems will be experienced at the commissioning phase, and could theoretically be attributed to the cabling, the LAN devices (such as hubs and switches), the Network Interface Cards (NICs) or the protocol stack configuration on the hosts.

The wiring system should be installed and commissioned by a certified installer. This effectively rules out wiring problems for new installations, although old installations could be suspect.

If the LAN devices such as hubs and switches are from reputable vendors, it is highly unlikely that they will malfunction in the beginning. Care should nevertheless be taken to ensure that intelligent (managed) hubs and switches are correctly set up.

NICs rarely fail and nine times out of ten the problem lies with a faulty setup or incorrect driver installation or an incorrect configuration of the higher level protocols such as IP.

5.16.10. TCP/IP

This section deals with problems related to the TCP/IP protocol suite. The TCP/IP protocols are implemented in software and cover the second (Internet), the third (Host to Host) and the upper (Application) layers of the ARPA model. These protocols need a network infrastructure as well as a medium in order to communicate. This infrastructure is typically Ethernet.

Typical network layer problems: If the TCP/IP protocol stack is not properly installed on the local host (host unable to access the network).

The easiest way to confirm this, apart from checking the network configuration via the control panel and visually confirming that TCP/IP is installed for the particular NIC used on the host, is to perform a loop–back test by pinging the host itself.

This is done by executing ping local host or ping 127.0.0.1. If a response is received, it means that the stack is correctly installed.

Other possible problems include:

- A host failing to obtain an automatically assigned IP address
- Reserved IP addresses
- Duplicate IP addresses
- Incorrect network ID – different netIDs on the same physical network
- Incorrect subnet mask
- Incorrect or absent default gateway(s)
- MAC address of a device not known to user
- IP address of a device not known to user
- Wrong IP address.

Transport layer problems: Without really getting into the detailed treatment of the TCP protocol, there are a few simple things that a relatively inexperienced user can check.

- No connection established
- Incorrect port number.

Troubleshooting Radio: When troubleshooting an existing system, it is worth checking on a few issues discussed earlier. These are as follows:

- Frequency selection
- Interference from other radio equipment
- Inter-modulation problems
- Incorrect path loss calculation
- Radio modems.

5.17. Industrial network security

Networking has been one of the greatest driving forces behind the growth of the computer industry. While low cost desktop computing brought the power of the digital age to millions of users, the real power of distributed computing has been unleashed by interconnecting the computers via networks, which made sharing of hardware and data resources possible.

Thus, network security involves three distinct aspects:

Confidentiality: ensuring that information is accessible only to those authorized to have access.

Integrity: safeguarding the accuracy and completeness of information and processing methods.

Availability: ensuring that authorized users have access to information and associated assets when required.

The goal of network security is to prevent an attack on the assets of the target system and in case it cannot be prevented, to minimize the undesirable consequences of a successful attack by early detection and counte-rmeasures.

5.17.1. Security in the context of Industrial automation systems

The use of computer based systems for industrial automation is now commonplace. These can be broadly divided under the following classifications:

- Automation systems such as Programmable Logic Controllers (PLCs), several of which are networked to form an industrial automation network. A Distributed Control System (DCS) is a higher-end industrial automation network used for the control of more complex, special purpose equipment and processes. This often uses proprietary hardware and software, unlike a PLC based network.
- Supervisory Control And Data Acquisition (SCADA) systems, which collect data from geographically dispersed resources and allow remote monitoring and control usually used in utility systems such a electric power and water supply.

5.17.2. Network security solutions

Network security threats are countered using the following approaches.

- Authentication, Authorization and Accounting (AAA)
- Encryption of data
- Access control, boundary routers, firewalls and filtering
- Intrusion detection and response.

Two other technologies need to be mentioned in this regard. One is the Virtual LAN (VLAN) is used to reduce the Internal security violations to provide a degree of control not usual in a normal LAN and the other is the Virtual Private Network (VPN). Security was not, however, the primary objective of the VLAN; it was rather the need to reduce congestion of the networks.

5.18. Network threats, vulnerabilities and risks

The goal of network security is to prevent an attack on the assets of the target system from succeeding and in case it cannot be prevented, to minimize the risks due to undesirable consequences.

“Risk is an expression of the likelihood that a defined threat will exploit a specific vulnerability of a particular attractive target or combination of targets to cause a given set of consequences.” It is not just anyone who can pose a real and serious threat to a network. The person should have adequate technical knowledge of how systems operate and the possible vulnerabilities that can be exploited, and should have adequate motivation to mount an attack, especially with the knowledge that it is a criminal act and carries substantial penalties.

So we have:

- Threats
- System knowledge
- Motivation
- Vulnerabilities which these threats employ to attack the target assets
- Consequences of attacks.

Network security would have achieved its goal if it minimizes the risk of undesirable consequences due to an attack. Threats from those with adequate system knowledge and the motivation to mount an attack exist. The vulnerabilities are also real. That an attack will happen is only a matter of time and should be considered a certainty. The security measures should aim to prevent the attackers from penetrating the system, but in a situation where the system is breached it should detect the intrusion and take appropriate counter-measures to reduce or eliminate undesirable consequences.

The assets being protected can include many things, some of them tangible and the others not. Consider the list below:

- Industrial facilities
- Employees
- Financial resources
- Trade secrets
- Reputation.

The first three are examples of tangible assets. Trade secrets are essentially intellectual property, which can be stolen or destroyed, in their physical form. Reputation is an intangible asset that can be affected adversely by service disruption, loss of data, and substitution of incorrect data etc. Refer to Figure 5.16.

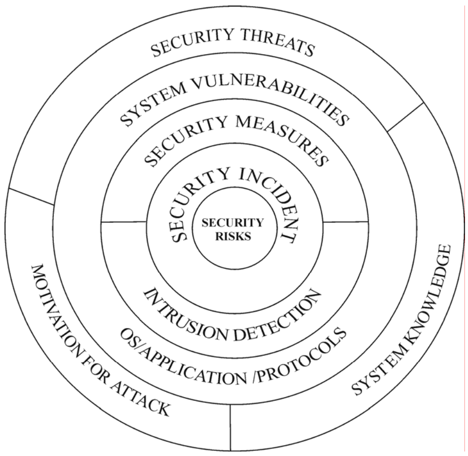


Figure 5.16
Security goal

In the outermost perimeter, we have threats, motivation and system knowledge. All these elements need to be present for an attack on a network. The attack is then started by studying the system for vulnerabilities (shown in the next layer).

A security incident happens when an attacker finds a vulnerability that can be used to break into the system. This is the next layer. The security incident has to be countered by the security measures (firewalls, encryption etc.) that deny the attacker an opportunity to get into the critical system areas, as well as other measures, which include the detection of an intrusion (the incident) and the response. If these measures fail, the attack becomes successful, opening up the system and the organization to the risks of security failure, (shown in the innermost circle). If these risks are anticipated and organizational measures are in

place to tackle them, the damage to the system or organization can be reversed or reduced and the effects of an attack thus minimized.

5.19. An approach to network security planning

Network security is not a matter of technology alone, but should focus instead on appropriate controls based on a clearly defined security policy. To determine these security policies, one needs to think about the business and examine the risks. You need to place a value-and a probability on them. You need to budget, to find the best way to spread the available money across the security options - and accept the unavoidable fact that it is not going to be perfect. You need to plan the implementation, and make sure that the rest of the organization (represented by its management) and the users understand and cooperate with the security measures that they are expected to follow. These principles are applicable not just to business networks as is commonly presumed, but to automation networks as well. Without a proper assessment of security needs and establishment of appropriate security policies, the best of security architecture, hardware and software may not protect the organizational information assets.

Connectivity to the Internet by different segments of the organization is quite essential from a business point of view. Even in Industrial Automation systems, such as SCADA networks of large utilities, Internet connectivity including access to corporate email services as a matter of necessity. This means that there is a need for connecting Industrial Automation networks to the business networks and then on to external organizations/services through the Internet. Also, remote access of the corporate network by users, either from homes or from remote locations, has become a matter of routine. All these needs, coupled with the inherent weaknesses in the technologies used make the network administrator’s task far more complex.

But threats are not just external. Many attacks take place from within the organization. A security system designed to protect data resources should therefore take both external and internal threats into consideration.

In fact, security of networks should not be looked at in isolation but should be a subject of a systematic study.

The following are the minimum steps required for arriving at a comprehensive network security system:

- Evaluate the risks in terms of tangible and intangible effects
- Plan for preventative measures
- Provide for detection of an attack and response
- Plan for the recovery of systems (in the event of a successful attack)
- Prepare a security policy document
- Arrange for dissemination and implementation of the policy
- Provide guidelines for auditing and monitoring of security measures (Including periodic review of the security policy itself)

5.20. Securing a network by access control

The Internet has become an important business enabler. Threats are thus faced by organizations not only from insiders (those who operate from within the local network) but also from outsiders accessing an organization’s resources through the

Internet. One of the devices that networks use for preventing unauthorized access is the firewall. Since the Internet is an untrusted network, the information resources of an organization have to be protected by providing security at the point of connection to the Internet using a 'perimeter router' - the simplest form of firewall. The perimeter router (also called a boundary router or edge router) provides protection using what is known as an access control list or access list. Firewalls are also provided to separate the internal as well as external users from important network assets such as the application servers and other servers providing FTP services, email services etc.

ACL: An ACL (Access Control List) is essentially a list of statements that filters unwanted packets by restricting network use by certain users and devices. ACLs can be used to block packets from specified source addresses, packets bound for specified destination addresses or to indicate that a packet is carrying information of specific interest.

Firewall; A firewall is either a software program or a hardware device that filters the information coming into a private corporate network, a specific part of the network or a personal computer connected through a modem. Using a certain pre-determined set of rules, a firewall acts as a filter for incoming packets of information. If the filters flag a packet, it is not allowed through.

Firewalls protect sites from attacks by outsiders who use the inherent vulnerabilities in the TCP/IP protocol suite. Additionally, they help mitigate security problems associated with an insecure system and with the inherent problems in providing a robust system of security for a large network of computers. There are several types of firewalls, from perimeter routers that can provide access control, to more powerful firewalls that can protect against vulnerabilities in the TCP/IP protocol, to even more powerful firewalls that can filter packets based on the content of the traffic. Usually large corporations install firewalls both for security against malicious incoming traffic as well as to filter out access to inappropriate sites by its internal users.

5.21. Authentication, Authorization, Accounting & encryption

AAA (Authentication, Authorizing and Accounting) and encryption of data are two of the main components in the security scheme of any network.

The security of any network depends on the control of who can access it and for what purpose. 'Authentication', the first component of AAA, achieves this objective of validating the identity of a user before permitting access. It is equally, if not more applicable to users who access the network from a remote location using public communication media such as the Internet.

The second component, 'authorization', determines which specific services and resources the authenticated user can have access to. Authorization defines attributes and privileges of resources, which each user is authorized to access and the activities that can be legitimately performed. This data can be stored locally on the network or in a centralized remote server if it is convenient to do so. Authorization thus provides a method for remotely controlling access where needed.

5.21.1. Use of the remote security database

The security information is stored in local security servers forming part of the network being accessed. This information includes usernames and passwords for all the network hosts and other devices such as routers. With a very large number of users, multiple network access servers become necessary. Instead of duplicating the security database in all of these servers, centralized information is held in a remote host which, besides the security database, will also hold authorization and accounting information.

A remote security database can make the management of Network Access servers simpler. It consistently enforces security policies to dial up users and manages the required accounting functions.

5.21.2. Encryption

Encryption is an important component of data security. It is constantly stressed that sensitive information, such as usernames and passwords used for authentication, is not to be sent in clear text. One way of sending this information is to instead send a value calculated by a hash algorithm. The other method is to encrypt the data.

All organizations are wary of sending their data over untrusted or public networks since there is always a possibility of the data being captured and read or modified.

Encryption helps in:

- Maintaining data integrity
- Maintaining privacy
- Ensuring that the data is authentic

Encryption refers to the deliberate alteration of data using a key (which is a fixed length string of bits) so that the data is meaningless to anyone intercepting the message unless he has the key. The reverse process of getting the data back from the encrypted form is called decryption and is done by the receiver using the same key. It is essential that both sender and receiver share the key. This method, where a shared key is used by both sender and receiver, is called symmetrical encryption or Private Key encryption. It is possible to attack an encrypted message (an attempt to decode it) using the 'brute force' method through an automated process of trying out all possible key combinations. Therefore the longer the key length, the more difficult it is to break. A 128 bit string will take thousands of years of computational time to break by 'brute force' method and is therefore considered safe. However, a 64 bit key can easily be cracked with current technology.

In addition to user authentication, Certification Authorities (CA) also provide digital certificates containing the public key. An enterprise can implement its own certification using certification servers or use third party services such as Verisign for issue and sharing of their public keys. This process of managing public keys is known as the Public Key Infrastructure (PKI). The PKI uses the hash functions, shared keys or public and private keys. The use of the PKI is important in situations that require a non-repudiation feature.

5.22. Intrusion detection systems

An intruder is one who attempts to gain unauthorized access to a network. Once he is in, an intruder can manipulate data, misuse the network resources and disrupt network services. An intruder detection system can identify an intrusion and send alerts to specified users as it is happening so that necessary measures can be taken.

An Intrusion Detection System (IDS) is not a substitute for other security measures such as proper AAA implementation, encryption of data or firewalls, but merely a reinforcement of these measures. In the event of the network security devices failing to stop an attack for whatever reason, an IDS will act as a back-up measure to detect the attack taking place and initiate a suitable response.

Intrusions happen due to a variety of reasons. They are:

- The absence of proper network policies
- Improper system configuration
- Technology weaknesses.

IDS systems can therefore be classified under the following two categories:

- Network-based systems
- Host-based systems

Network-based IDSs: Network-based IDSs monitor the network packets flowing through a specific section of the network to detect an intrusion. They deploy a network adapter operating in promiscuous mode, which means that they read and process all packets regardless of the destination address.

Host-based systems: In a host-based IDS system, detection agents are deployed on all computers and report intrusions to a managing agent installed on a central computer. The detection agents operate by sharing the disk and memory available in the computers, which may cause a degradation of performance. Host-based IDSs are not suitable for large networks. These systems are best suited where there is a need to constantly monitor specific hosts.

A response to an attack can be either active or passive. A passive response is one where the IDS simply generates an alert and leaves it to system personnel to intervene and take action. Usually, an alarm is by means of a pop-up window on the administrative console.

5.23. VLANs

When a LAN is divided into segments using a switch, with each port serving a smaller number of network nodes, the chances of collision reduces. Moreover, the devices that normally communicate with one another are placed in one segment so that the need for forwarding the packets to other ports also gets reduced. In some cases, machines that require very high bandwidth (for example, a server or a high performance workstation) are connected directly to a switch port, thus enabling them to have almost the entire bandwidth of one segment dedicated to them.

The need for VLANs: Very often the personnel involved in a particular project or those belonging to a particular department are not confined to a given area and are spread throughout a building or campus. Product design teams may be cross

functional groups and usually exist for short periods of time. In such cases, grouping the users into one physical segment is not feasible. In these cases, more packets have to travel from one physical segment (or switch port) to another, thus increasing the network loading. VLANs offer a way to overcome these problems.

A VLAN logically groups switch ports into workgroups. Since broadcasts and multicasts between the users of a workgroup are likely to be high, a VLAN limits the broadcast traffic to within the particular virtual network and thus performs like a virtual broadcast domain.

Benefits of a VLAN: VLANs offer a number of advantages over the traditional LAN implementation:

- Performance improvement
- Improved security
- Ability to set up virtual workgroups
- Reduced administration
- Reduced cost.

5.24. VPNs and their security

A VPN is basically a corporate network that is built around the communication infrastructure of the Internet rather than using leased lines or a Remote Access Server using direct dial-in. Since the Internet is a public medium where the traffic is prone to interception or modification, unlike the privacy offered by dedicated leased circuits, security issues play an important role in the implementation of a VPN. A VPN is however a highly cost effective proposition, as dedicated lines are required only to connect the corporate network to an ISP (usually located within the same city).

5.24.1. Types of VPN

VPN solutions are essentially of three distinct types:

- Inter-site or inter-LAN VPNs
- Remote access VPNs
- Extranets

While all the three of these types of connectivity are essential from the enterprise viewpoint, most of the savings result from Remote Access VPN. This is because:

- Cost of remote access and the number of employees who travel and need to connect using long distance dial up are showing an increasing trend
- A dial-up Internet connection offers good bandwidth and is therefore becoming acceptable to more users, particularly those using applications based on client server technology and multi-tier architectures that conserve bandwidth
- A local dialup connection using a reliable Internet Service Provider (ISP) offers a very high degree of availability and Quality Of Service (QOS) level compared to direct dial up through long distance lines.

5.24.2. Requirements for designing a VPN system

Any enterprise planning to implement a VPN system must carefully evaluate the various issues of importance. A 5-tier model proposed by the Gartner Group sums up these issues and can be a starting point. See Figure 5.17.

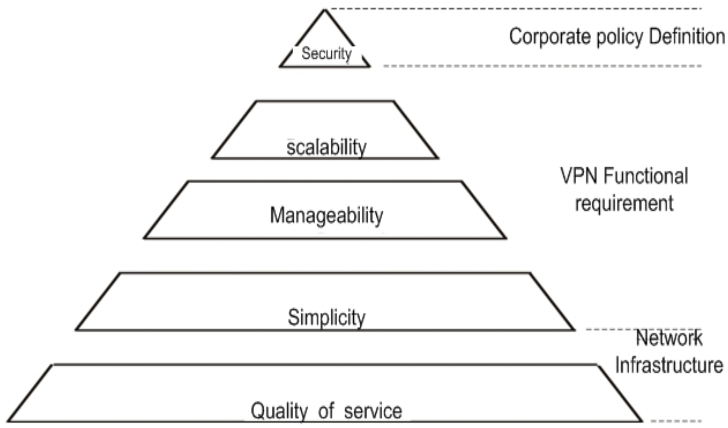


Figure 5.17
A 5-Tier model for VPN implementation

The 5 tiers are: security, scalability, manageability, simplicity and quality of service. Security is a factor decided by the corporate policy. Scalability, manageability and simplicity are functional requirements and will depend on present and perceived future needs, particularly the issue of scalability. Quality of service will be primarily dependant on the ISP whose infrastructure will be used for the VPN.

5.25. Wireless networks and their security issues

Wireless technologies, in the simplest sense, enable two or more devices to communicate without physical connections. Wireless networks serve as the transport mechanism between devices, among devices and the traditionally wired networks (such as Enterprise networks and the Internet). Wireless networks are frequently categorized into three groups based on their coverage range

WLANs allow greater flexibility and portability than do traditionally wired LANs. Unlike a traditional LAN, which requires a wire to connect a user's computer to the network, a WLAN connects computers and other components to the network using an Access Point (AP), which connects to the wired Ethernet LAN via an RJ-45 port. APs typically have coverage areas of up to 300 feet (approximately 100 meters), referred to as cells. Users move freely within the cell with their laptop or other portable network devices. Access points can be interlinked to allow users to even roam within a building or between buildings.

5.25.1. Security risks

Risks in wireless networks are equal to the sum of the risk of operating a wired network plus the new risks introduced by weaknesses in wireless protocols. To mitigate these risks, agencies need to adopt security measures and practices that help bring their risks to a manageable level. They need, for example, to perform security assessments prior to implementation to determine the specific threats and

vulnerabilities that wireless networks will introduce into their environments. In performing the assessment, they should consider existing security policies, known threats and vulnerabilities, legislation and regulations, safety, reliability, system performance, the life-cycle costs of security measures and technical requirements. Once the risk assessment is complete, the network security administrator can begin planning and implementing the measures that will be put in place to safeguard the systems and lower the security risks to a manageable level. The security administrator should periodically reassess the policies and measures in place because computer technologies and malicious threats are continually changing.

Chapter 6. HAZOPs Hazard Operations

6.1. Introduction

HAZOP stands for Hazard and Operability. It is a method of study for identifying hazards and operability problems in many operational situations ranging from chemicals and fuels processing through to electrical and machinery systems. It is also finding applications in the planning of operational activities such as emergency response and disaster management.

The HAZOP technique is applied to piping and instrumentation drawings (P&IDs) of a process.

For a new facility, ideally a HAZOP workshop would be carried out at the end of the detailed design stage when all P&IDs have been finalized and issued for detailed design and construction.

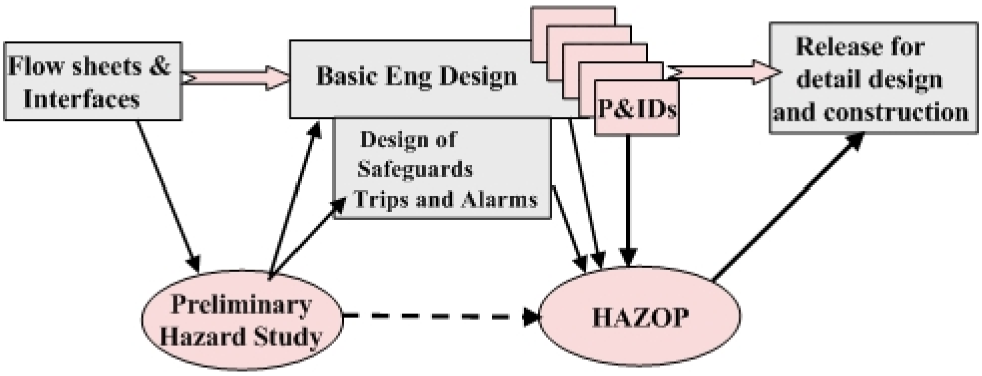


Figure 6.1
HAZOPS should be done when the design is ready for detail engineering

Therefore, in practical terms, the HAZOP workshop is carried out as late as possible in the detailed design stage when P&IDs are as complete as possible while still allowing sufficient time for HAZOP recommendations to be considered and incorporated into the design.

6.2. HAZOP Workshop

In preparation for the HAZOP workshop, the facilitator will have agreed with the client/project on:

- Terms of reference
- Documents required
- Team membership and availability
  - Team Structure
  - Team attitudes
- Estimated duration of workshop
- Venue for workshop meetings
- Reporting method
- Scope of work
- Timing of HAZOP Studies

6.2.1. Team Leader and Team

The first requirement of the team leader is to see that the Hazop methodology is used effectively and productively. The IEC 61882 standard describes the study leader’s skills and duties as follows. He should:

- Not be closely associated with the design team and the project
- Be trained and experienced in leading HAZOP studies
- Be responsible for communications between project management and the HAZOP team
- Plan the study
- Agree to the study team composition
- Ensure that the study team is supplied with a design representation package
- Suggest guidewords and guideword-element/characteristic interpretations to be used in the study.
- Conduct the study
- Ensure the results are documented

Who is in the team?

In addition to the team leader the essential players are:

- Recorder or “scribe”.
- Designer (process engineer, control engineer, mechanical engineer etc according to project).
- Project engineer (may also be designer).
- User (commissioning manager or production manager).
- Instrument/control engineer.
- SHE expert (mandatory in some countries).
- Contractor and Client representatives.

6.2.2. Good HAZOP Workshop Records

- Record during the HAZOP Workshop
- Record comprehensively
- Record format and content.

6.2.3. Quality HAZOP Reports

Reporting: Depending upon the client/project, the required format of the report may vary.

Contents of the Report: The contents list is based on descriptions in IEC 61882 and EPSC guide.

6.2.4. Hazard Identification and Risk Management

Hazard analysis is used to help quantify the risks associated with a hazard. The task of Hazard analysis includes.

- Estimating how often an incident (Hazardous event) will occur.
- Estimating the consequences to persons, environment and plant.
- Deciding on the required amount of risk reduction (if any).

Two methods of hazard analysis are widely used.

Failure mode and effects analysis (FMEA): looks at possible component faults and tabulates their impact on risks.

Fault tree analysis: looks at a hazard event and resolves the causes into basic events.

6.2.5. Cost Consideration

Market Structure: Hazard analysis & cost consideration rules mandated across an industry will have different impacts on the industry, depending on the market structure of the industry.

Policy Issues: Additional issues that must be considered for a complete evaluation of the benefits and costs of food safety risk reductions relate to information, public versus private intervention, accurateness of illness estimates, marginal benefit-cost analysis and efficiency in production.

Chapter 7. Safety Instrumentation and Machinery

7.1. Introduction

Safety instrumentation is not exclusively an instrument and control engineering subject. The successful implementation of a safety system project depends on the support and knowledge of other disciplines as well as being dependent on a full commitment from company management structures. It requires the environment of a well defined safety management system within the company. Without proper support structures and a good understanding by all involved in defining safety requirements the safety instrumentation on its own will be unlikely to deliver the levels of safety that are expected of it.

The support structures are a crucial part of the assessment scope for compliance with the new IEC 61508 and 61511 standards. It is the responsibility of the instrument engineer to involve colleagues from other disciplines in the safety package. It is the responsibility of the management to see that the safety activities are clearly assigned and supported.

7.1.1. Safety system basics

7.1.1.1. Definition of Safety Instrumented Systems

Safety Instrument Systems are control systems that take the process to a safe state in terms of conditions that may be hazardous or could eventually give rise to a hazard if no action were taken. They perform “safety instrumented functions” by acting to prevent the hazard or mitigating the consequences.

The abbreviation SIS is used for “Safety Instrumented Systems” while the abbreviation SIF means “Safety Instrumented Function” which is the task or function performed by the SIS. These are terms generally used in engineering standards. Other names may be used, however, because of the different ways in which these systems have been applied.

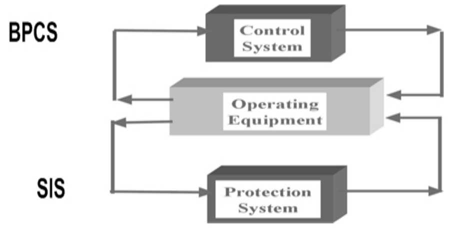


Figure 7.1
SIS operates independently of the Basic Process Control System (BPCS)

The SIS is an example of a "Functional Safety System." In other words safety depends on the correct functions being performed. This distinguishes functional safety from "passive safety "devices such as handrails, or blast proof walls. It is a useful term because it distinguishes the active safety system, in its range of uses, as a system that must function properly to provide safety.

7.1.1.2. The structure of an SIS

Safety Instrumented Systems are normally structured into three parts within a framework or boundary that defines it:

- Sensor sub-system: To capture the data on line from the process
- Logic solver sub-system: To evaluate the data and make decisions on when and how to act
- Actuator sub-system: To execute the required actions on a plant

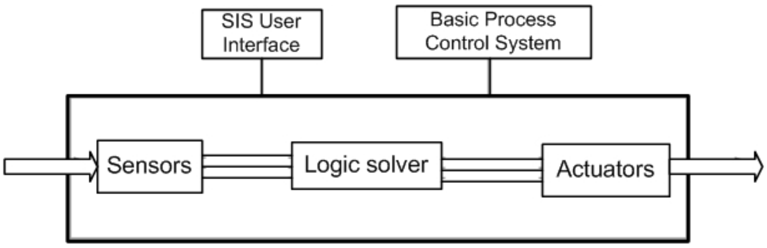


Figure 7.2
Structure of a Safety Instrumented System

Figure 7. shows that the subsystems lie within a boundary that defines the essential SIS while it also needs to have interfaces to its users and those who maintain it as well as to the basic plant controls. Items within the boundary must be engineered to the standards required for functional safety systems.

All three sub-systems must perform correctly to ensure that the SIS can provide the required protection, which brings us to one of the key design principles.

7.1.2. Risk reduction and safety integrity

There is a common saying in the control systems world: "if you want to control something, first make sure you can measure it." We need to control the risks of harm or losses in the workplace due to hazards of all forms. Therefore, Risk is to be measured before controlling it.

7.1.2.1. Measurement of risk

Risk can be evaluated qualitatively or quantitatively. The qualitative approach requires that risk is described in such terms as "high" or "low' or "moderate". These terms are only effective if everyone has a good understanding of what they mean in the context of use. Hence a "high risk neighborhood" is not popular with insurance companies. If the terms are well defined or "calibrated" against a scale of values that is generally accepted the qualitative risk measurement can be very effective.

The quantitative approach is easier to define in terms of frequency of events and then the number of people getting hurt, but it is often hard to extract a firm number from a situation without a lot of statistical evidence.

7.1.2.2. Safety integrity

The degree of confidence that can be placed in the reliability of the SIS to perform its intended safety function is known as its "Safety Integrity". The concept of safety integrity includes all aspects of a safety system that are needed to ensure it does the job it is intended to perform. One of these aspects will be the hardware reliability of the equipment and the way it responds under all conditions. Other aspects include the accuracy with which it has been designed and the level of understanding of the hazards that went into its original design

7.1.3. Protection layers

Now that we see the SIS as a risk reduction element, it is helpful to see how it fits into the context of overall plant safety. This will enable us to see how the SIL target can be adjusted to provide best overall value from the plant safety systems.

Belt and braces

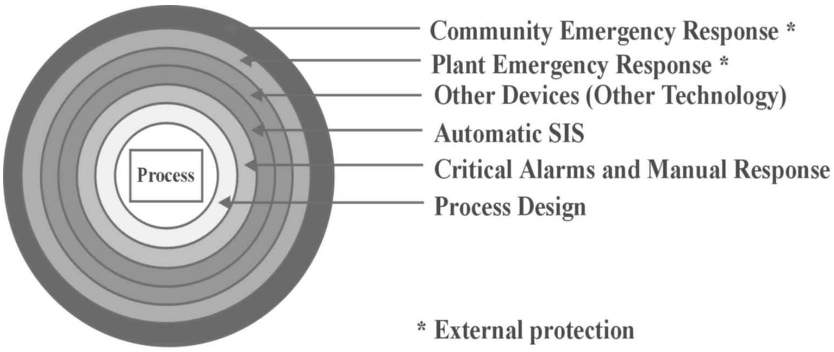


Figure 7. 3
Layers of protection model

The concept of protection layers applies to the use of a number of safety measures all designed to prevent the accidents that are seen to be possible. Essentially, this concept identifies "belts and braces" involved in providing protection against a hazardous event or in reducing its consequences. Figure 7. shows the concept where the core risk, due to a hazard, is seen to be contained by successive layers of protection leaving a minimal or acceptable risk level at the outside boundary.

Protection layers can be divided into two main types:

- **Prevention layers**: These try to stop the hazardous event from occurring
- **Mitigation layers**: Mitigation layers reduce the consequences after the hazardous event has taken place

To summarize, SIS is just one component of an overall risk management strategy for a hazardous activity in a manufacturing plant. For a SIS to be effectively designed and implemented, the following key aspects of a SIS project will have to be assured.

- Identify the hazards and estimate the risks: Hazard studies and hazard analysis
- Define the overall safety targets for each type of risk: The overall amount of risk reduction needed for the hazard needs to be defined by someone who knows what is acceptable: This is a management or corporate responsibility
- Allocate risk reduction functions and RRF's to layers of protection: This defines the risk reduction contribution of the SIS and hence defines its target SIL
- Ensure that each safety layer is managed to deliver the required risk reduction: This requires correct design procedures in each discipline and requires work procedures and responsibilities to be defined and supported by management
- Ensure that the SIS delivers the required functional safety

7.1.4. Safety management principles

Safety management principles help to look at the principles of risk management because they can be applied directly to safety management. Understanding risk management will show us how the application of Safety Instrumented Systems is an integral part of the overall task of managing risk in a company.

The meaning of safety management: Safety management involves the provision of a safe working environment for all persons involved in the manufacturing process. It extends to cover the safety of the environment and the security of the business from losses.

The fundamental components of safety management will include:

- Having a systematic method of identifying and recording all hazards and risks presented by the subject plant or equipment
- Ensuring that all unacceptable risks are reduced to an acceptably low level by recognized and controllable methods that can be sustained throughout the life cycle of the plant
- Having a monitoring and review system in place that monitors implementation and performance of all safety measures
- Ensuring all departments and personnel involved in safety administration are aware of their individual responsibilities

- Responding to regulatory requirements from national and local authorities for the provision of adequate safeguards against harm to persons and the environment.

Maintain a risk register and a safety case report that demonstrates adequate safety measures are in place and being maintained at all times. Safety management is effectively the same as the more general term, risk management, but applied specifically to risks associated with harm to persons, property or environment.

Risk management: Risk management is a very broadly used term and is typically applied to business and organizational activities.

Managing risk

- Requires rigorous thinking. It is a logical process, which can be used when making decisions to improve the effectiveness and efficiency of performance.
- Encourages an organization to manage pro-actively rather than reactively.
- Requires responsible thinking and improves the accountability in decision making.
- Requires balanced thinking "Recognizing that a risk-free environment is uneconomic (if not impossible) to achieve, a decision is needed to decide what level of risk is acceptable".

Requires understanding of business operations carried on, where conformity with process will alleviate or reduce risk.

Conclusions from risk management: We have seen how the generalized models for risk management are directly applicable in safety management. Risk management involves the systematic analysis of risk levels, knowledge of acceptable risk levels and the selection of measures to reduce risk to the acceptable level. The selection of measures involves balancing the level of safety achieved against the cost of achieving it.

When we look at the new application standards for Safety Instrumented Systems it is easy to recognize the same principles being applied. Industry therefore has available a set of recognized standards and practices for designing and operating safety systems that aligns with well established principles of risk management.

7.1.5. The legal framework for process safety

Most industrialized countries have legal frameworks in place that are similar in nature and have been substantially improved in recent years. Safety regulation now emphasizes the need for a complete safety management system. This aims to deal with the fact that many accidents can be traced back to failures to manage the various aspects of safety from identification of hazards through to training and continued monitoring of safety performance.

These provide a good indication of what one should expect to be doing to satisfy good practices anywhere. The most commonly seen principle is that all potentially hazardous activities must be subject to a risk assessment process. This comprises:

- Hazard studies to identify hazards and risks
- Risk analysis to decide the level of risk
- Risk reduction measures to decide if they are needed
- Risk reduction measures to be implemented
- Confirmation that the process is now safe to an acceptable level of risk
- Periodic audits and reviews of safety studies to be carried out

In the case of process industries, plants having a known hazardous process or having major accident potential are required to develop a comprehensive safety case for inspection by authorities. This includes proving that they have a good safety management system in place. They are required to carry out process hazard analysis studies at frequent intervals to ensure the plant risk assessments and treatment methods are up to date with the current version of the plant.

7.1.6. New standards

In this section, we look at the background to the new standards and examine some of the key features that make the standards of great value to system users and designers.

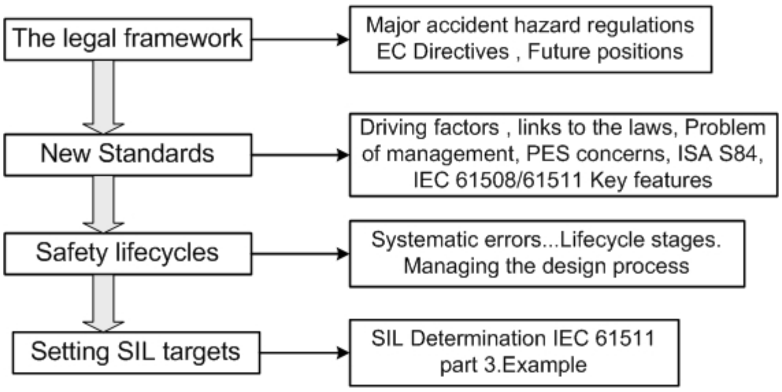


Figure 7.4
Section of roadmap

Two relevant standards exist for SIS practices. Which one is to be used?

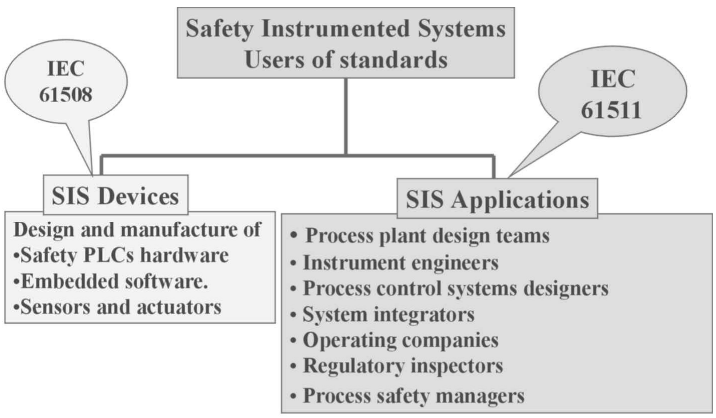


Figure 7.5
Users of the IEC Functional safety standards for SIS

IEC 61511 is to be used by those who are managing, designing, implementing or operating a safety instrumented system application in a process or similar plant. The safety equipment products bought from a system supplier or instrument vendor should be engineered in accordance with IEC 61508. IEC 61511 should be used for plant safety projects and IEC 61508 for design and manufacture of safety system products.

IEC 61508: The standard emphasises the life cycle approach to the overall safety system project. Perhaps the most significant feature to note is that conformity to this standard requires both technical items and the overall management of the safety project to be in compliance with the mandatory parts of the standard.

The scope includes all project stages from initial concepts and hazard studies through to operation, maintenance and modification. The standard covers electrical, electronic and programmmable electronic systems and lays down standards of engineering and quality assurance for both hardware and software.

The standard is large and requires a lot of study for effective use in a company. However it is of great value to both manufacturers and end users and despite some criticisms of its complexity, it has been adopted by many large processing companies as their standard reference for design and operation of Safety Instrumented Systems.

7.1.7. The safety lifecycle

It is important to examine why the standards advocate a safety life cycle approach to an SIS project and see how this approach shows the way to plan and manage a safety project.

The safety life cycle is an orderly sequence of design and builds stages with each activity being mapped out with essential requirements that must be satisfied at each phase. It is visualized by a flow chart showing the procedures suggested for the management of the safety functions at each stage of the life cycle. Figure 7. shows a simplified version that identifies the main stages.

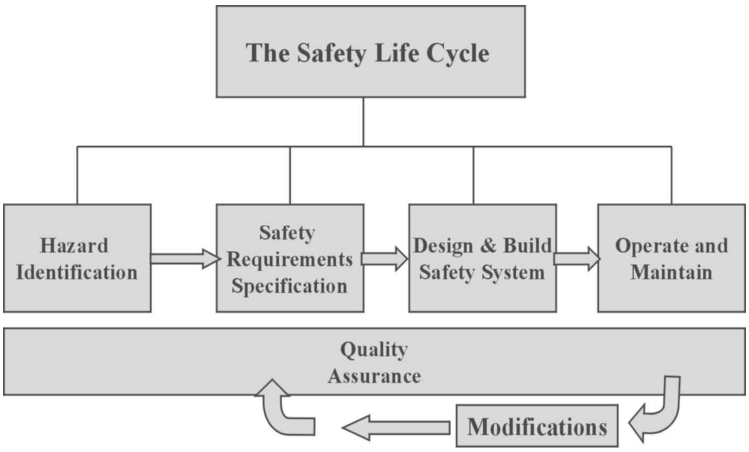


Figure 7.6
A simplified version of the SLC that identifies the main stages

Safety projects begin with a phase where the scope of the plant is defined and the hazards are identified. The hazard study identifies hazards and a hazard analysis leads to the estimation or ranking of risks.

The next stage involves determining the risk reduction needed to meet a safety target and this leads to the production of the Safety Requirements Specification (SRS); a key document that will be a design reference throughout the life of the safety system.

The design and building stages of the SIS project can now follow from the approved baseline of the SRS. As this stage progresses, the project team will continue to carry out verification exercises to ensure that the design remains true to the SRS and that the SRS remains true to the hazard studies.

Before the SIS enters service, it will be subjected to a rigorous validation exercise that involves testing of the functions against the details defined in the SRS. It then moves into the operations and maintenance phase where a strict set of procedural rules will be implemented.

Change management procedures remain in force throughout the safety lifecycle and any recorded changes are recycled back through the relevant design and testing stages. This ensures that all documents and decisions remain current for the present state of the plant.

7.1.8. Introduction to safety PLCs

Safety PLCs have become the dominant form of logic solver in the past 10 years through their ability to provide shared logic solver duties for many safety functions within one SIS. They offer the facilities needed by most safety functions to perform fairly simple logic combined with efficient operator interfacing and secure management of the program logic.

Safety PLCs are specially developed for their tasks through the provision of extensive diagnostic coverage using internal testing signals operating between scanning cycles of the application logic. Effectively, the PLC detects its own faults and switches itself into a safe condition before the process gets into a dangerous condition.

The software of a safety PLC is specially developed to have a range of error detecting and monitoring measures to provide assurance at all times that the program modules are operating correctly. The application programs are developed with aid of function block or ladder logic languages where each function has been extensively tested for robustness and only limited configuration options are available.

One major objection to safety PLCs has been their cost and this is particularly a problem for small plant applications. This is gradually being addressed as smaller and cheaper units are now available.

7.1.9. The cost of ownership

Having discussed something of the project activities and some technical aspects of safety system, it may be helpful to consider the issues of cost and justification for installing an SIS.

The justification for installing an SIS may be for one or more of the following reasons:

- It is essential for safety where no alternative methods exist
- It is the lowest cost option for safety
- It helps prevent environmental harm and/or guards against emission limit violation
- It helps protect against asset losses through plant damage and lost production capacity.

From the management perspective, it will be essential to have a measure of the costs involved in buying, maintaining and operating a safety system. If the true operating costs of the SIS can be evaluated, these will help to identify potential for cost savings and performance improvements. It will help to have an approximate cost model that can be used as a basis for establishing the total costs involved in SIS. With this in mind, it is possible to show how the case for performance improvement may be justified by further reducing operating costs.

7.2. Introduction to IEC 61511 and the safety lifecycle

The idea of a safety lifecycle model is to plan and control the various activities of a project so that each step follows logically and accurately from the previous step. The main steps are:

- Identification of plant scope and its hazards
- Evaluation of any risks to determine risk reduction needs
- Allocation of risk reduction duties to SIS and non-SIS layers of protection
- Development of the safety requirements specification
- The building and testing of the SIS to specification (known as the realization phase)
- Installation and testing of the SIS
- Operation and maintenance of the SIS
- The managing of changes to the SIS design or equipment

Each phase of the lifecycle requires input information and delivers output information. The relevant clauses of the standards define the inputs, activities and outputs for each phase.

IEC 61508 offers a safety lifecycle model that will serve any project and many companies may elect to use this version for their applications. IEC 61511 offers a similar project model but it has been designed specifically for process applications.

7.2.1. IEC 61508 SLC version

IEC 61508 version of the SLC is the most general version and forms the basis of all the IEC standards. Box numbers are used to reference a detailed set of clauses defining the requirements of the standard for that activity.

The clauses are easy to follow because they are defined in terms of:

- Scope
- Objectives

- Requirements
- Inputs from previous boxes
- Outputs to next boxes

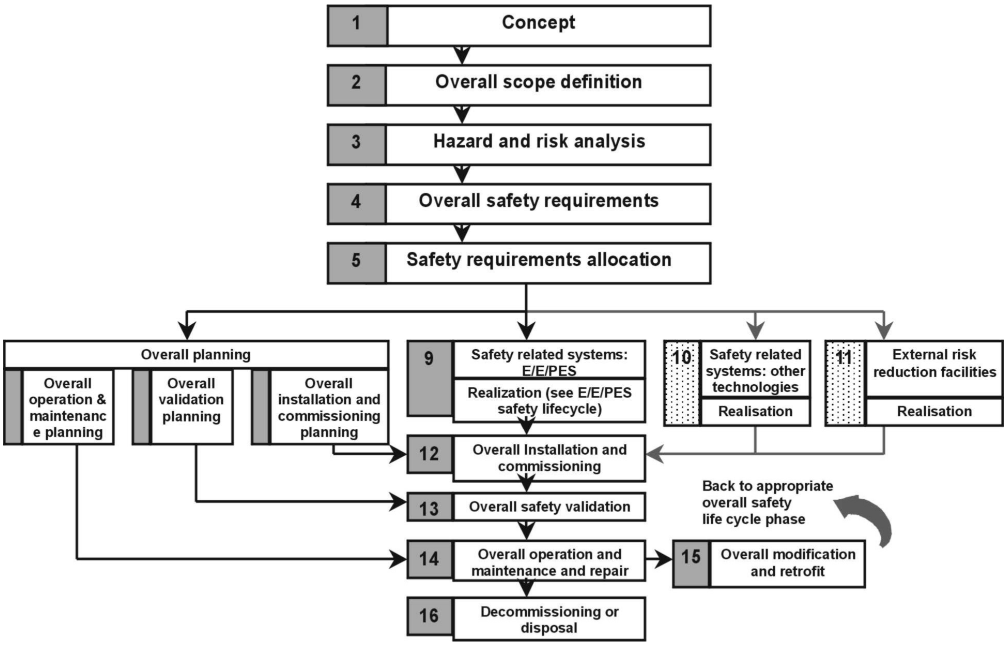


Figure 7.7
IEC 61508 SLC version

Developing the overall safety requirements: The first 4 phases deal with the tasks of defining the scope of the plant, identifying hazards and risks and deciding the overall safety requirements. This work has been defined as an integral part of the standard as it must be done according to the correct procedures to achieve compliance.

Safety allocations: Once the overall safety targets have been established, the SLC moves on to the "Safety Allocations" phase where the various layers of protection are defined and allocated a certain portion of the risk reduction task. This results in the SIS risk reduction task being clearly identified and hence the SIL targets can be defined for each individual safety function.

Realization phase: These stages are followed by the "realization" phase. This term describes the job of actually building the safety system and implementing any software that it contains. Large sections of IEC 61508 are concerned with the details of the realization phase and there are whole lifecycle models for the activities contained within this stage.

Validation and operations: Once the SIS has been built, the lifecycle activities move on to the "installation, commissioning and validation". All the standards place great emphasis on validation. This activity is seen particularly in the form of the final site acceptance testing using methods that test the SIS response under plant operating conditions in the most realistic way possible.

Finally we get to use the safety system for real duties and arrive at the operating and maintenance phase.

7.2.2. IEC 61511 SLC Version

Shown in Figure 7. is a version of the SLC model. Comparing this version with the 61508 model, we can see that the tasks have been grouped into more familiar sets of activities that will match up easily to the natural progression of a process safety project.

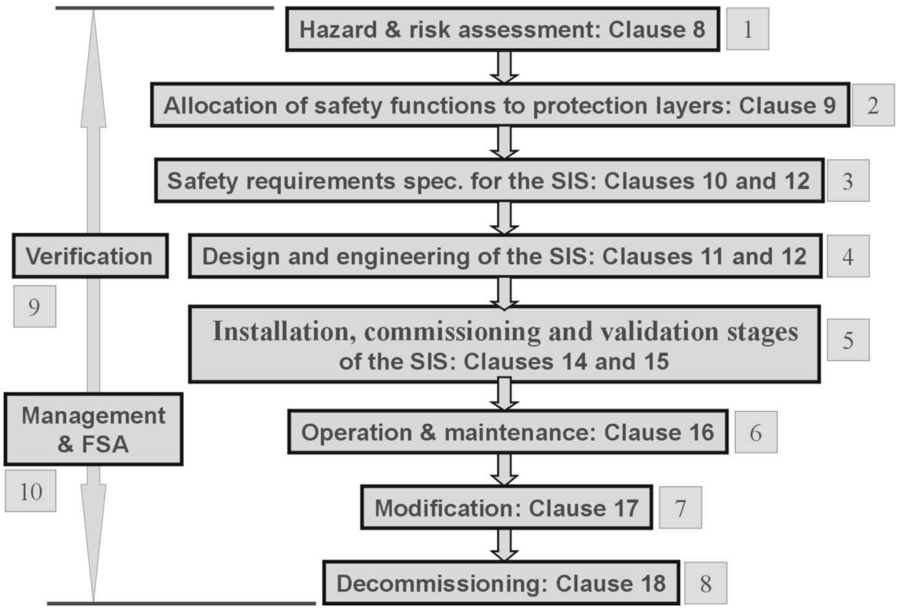


Figure 7. 8
IEC 61511-safety lifecycle model

Phase 1: Hazard and risk assessment

The model refers us to Clause 8. This describes objectives and requirements.

The first phase of the SLC delivers a sound basis of information about the hazards and records assumptions made about the risks. It is the essential foundation for the safety functions that will be needed.

IEC 61511 does not set out to provide detailed requirements for the hazard and risk assessment phase. It restricts itself to those aspects relevant to specifying the SIS requirements.

Phase 2: Allocation of safety functions to protection layers: clause 9

The idea of this phase is to decide on how much risk reduction is to be allocated to the identified or planned layers of protection.

The objectives of this phase are to:

- Allocate safety functions to protection layers
- Determine required safety instrumented functions (SIF)
- Determine the associated SIL for each SIF

The requirements clause requires us to identify all risk reduction measures and define each Safety Function (SIF) with its own SIL.

SIS safety requirements specification: clause 10

The Safety Requirements Specification (SRS) is a formalized and detailed document describing all essential functions of the SIS needed for the plant. This phase with the preceding two included is known in IEC 61511 as “Stage 1”.

SIS design and engineering: clauses 11 and 12

The requirements of this section comprise all the essential design constraints that the standard mandates. This is therefore the place to look for any design rules that are to be imposed on the SIS.

Clause 12 describes the requirements for the application software engineering and includes selection criteria for the utility software. This is the programming tools, compilers and display software that enables an engineer to configure the application logic using a high level language such function block or ladder logic. The specially restricted versions used in safety PLCs are described here as “Limited Variability Languages “or LVLs.

The end of this phase is achieved when all the design has been done, all the components and instruments have been decided and the software has been written, tested, integrated into the hardware platform and tested as a complete logic solver. Completion of the design and engineering activities is also known in the standard as “Stage 2”. Optionally this stage can include the Factory Acceptance Test (FAT).

SIS Installation, commissioning and validation: clauses 12.3,14, 15

This phase begins with the equipment at site and the logic solver FAT completed. Clause 12.3 is included here as this concerns safety validation planning.

After installation, Clause 14.2.3 outlines the commissioning requirements in terms of essential features that must be checked. The list includes things such as power supplies, removal of packaging, instruments calibration, instruments and logic solver operations and loops to be checked.

Clause 15 then describes the essential requirements for safety validation. In process control terminology this is the start up acceptance testing. Validation is of critical importance to safety system installations because it is the only way of knowing that the final result of the design and building effort can provide the required safety.

Completion of these activities is also known in the standard as “Stage 3”.

SIS operation and maintenance: clause 16

This phase covers the operating life of the SIS on the plant. Clause 16 of the standard defines the essential subjects for routine and abnormal operation of the SIS. The objectives are to ensure that the required SIL is maintained during Operation and Maintenance (O&M) and conversely to see that maintenance is adequate to keep the SIL at its intended level.

This stage is also known in the standard as “Stage 4”.

SIS modifications: clause 17

The objectives of this phase, defined in clause 17, are: "those modifications to any safety instrumented system are properly planned, reviewed and approved prior to making the change; and to ensure that the required safety integrity of the SIS is maintained despite any changes made to the SIS."

The completion of a modification activity is also known as "Stage 5"

SIS decommissioning: clause 18

This phase is similar to the modification phase because it requires an impact analysis on the effects on safety of de-commissioning.

Verification activities: clause 7 and 12.7

Clause 7 of IEC 61511 details the requirements for verification, which is essentially aimed at establishing that each phase has been completed properly and that the results are verified to be in accordance with the objectives of that phase and is traceable to the input information.

Assessment, auditing and revision: clause 5

Functional safety assessment and auditing are part of the overall requirements of IEC 61511 for management of functional safety. Clause 5 of IEC 61511 discusses these requirements. We have considered the organizational issues but when it comes to any of the project life cycle, the standard also requires that we carry out an assessment of how well the safety objectives have been met for that project.

7.3. SIS configurations for safety and availability targets

Once the requirement specification is established, the final design of the SIS can be carried out. The design process is concerned with finding suitable instruments in an independent SIS loop that has a structure or architecture appropriate to the SIL reliability targets. This also takes into account, the expected reliability and failure modes of the instruments. If the instrument is good, there is less need of redundancy for safety but this choice has to be made within a framework of rules.

7.3.1. The design process

To establish the design process, it helps to first look at what IEC 61511 requires to be done to build the SIS to meet the targets. Figure 7. spells out the evaluation process that should be followed once a design concept has been proposed. The preliminary evaluation of these points can be done at the concept stage and these points should be confirmed before the final design.

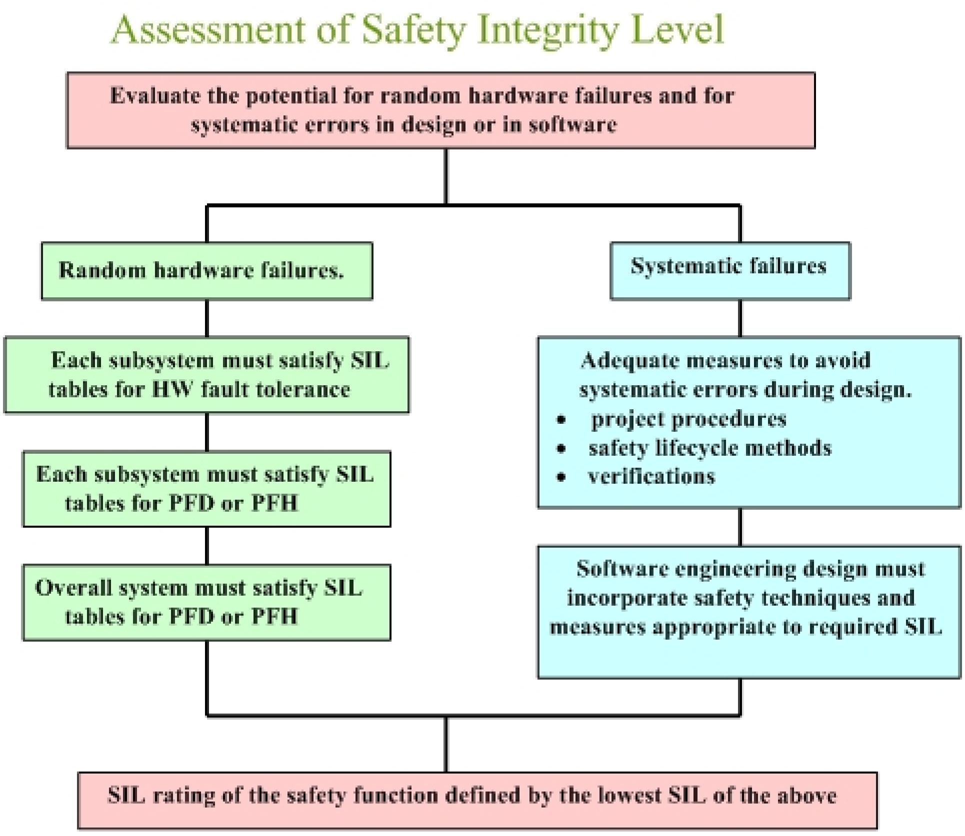


Figure 7.9
Overall design requirements for meeting the SIL target

Figure 7. shows that the assessment of the safety integrity level for an SIS design is divided into two measures a) that avoid systematic failures and b) that avoid random hardware failures. We will consider one of the ways in which hardware design can be developed to minimize the hardware failures. This introduces the subject of hardware fault tolerance.

7.3.2. Safety reliability versus availability for production

Functional safety is achieved by the SIS doing the safety job and when failures occur, the system must act in a manner that still ensures safety. i.e. the system will fail safely or will indicate a fault for suitable action. For a single channel SIS, this usually means shutting down the process even if it is only for the reason that the safety system has found a fault. This behavior meets the needs of the SIL target but it may not satisfy the business objectives of the process.

What is desired is a plant where safety is achieved at a reasonably economic cost. The components of cost here are:

- The capital cost and maintenance cost of the SIS
- The cost of accidents should they occur, including the production losses
- The lost production costs associated with achieving safety

The third item depends on whether or not the process operating costs are sensitive to downtime.

7.3.3. Architectures and fault tolerance concepts

Fault tolerance is one of the most important underlying principles of all safety systems whether used in chemical plants, machinery, automobiles or in business planning. If a safety system can still provide protection in the presence of dangerous faults, it reduces the chances that an accident can occur.

7.3.3.1. Terminologies

Dangerous failure: IEC 61511-1 defines dangerous failure as: "failure which has the potential to put the safety instrumented system into a hazardous or fail-to-danger state".

Dangerous detected failure: If the dangerous fault can be detected by some method of diagnostic testing or by the way the circuits are arranged, it will be known as a "revealed" or "overt" fault. When found by a diagnostic test procedure, these are usually known as "dangerous detected failures".

Safe failure: Conversely this is a failure that "does not have the potential to put the safety instrumented system into a hazardous or fail-to-danger state." It is also called "nuisance trip".

Safe detected failure: Some safe failures may not be revealed by causing a trip but they may be detected by other means such as a diagnostic test.

Common mode failure: Applies to any of the above failure modes where they are likely to affect two or more devices or instruments at the same time for the same reasons. Where redundant instruments of the same design or the same principles are used, the likelihood of common mode failures can be in the range 5% to 20 % of all failures. Without diverse instruments, common mode failures will limit the availability benefits achieved through redundancy.

Fault tolerance: Hardware fault tolerance is the ability of a system to continue to undertake the required safety function in the presence of one or more dangerous faults in hardware. Hence a fault tolerance level of 1 means that a single dangerous fault in the equipment will not prevent the system from performing its safety functions. A fault tolerance level of zero implies that the system cannot protect the process if a single dangerous fault occurs in the equipment.

Safe failure fraction: This parameter is vitally important for the safety integrity or confidence level and it will determine if greater levels of fault tolerance has to be built into a given application.

7.3.4. Identification of subsystems

The first step in evaluating the structures of an SIS is to recall that the SIS is conveniently divided into subsystems of Input, Logic and Actuator. See Figure 7.. These subsystems must be decided at the beginning of the design procedure. It is not always as simple as identifying the instruments and the valves or motors because the signal transmission and conversion devices will also fall into the sensor and actuator subsystem. What about the input and output circuits of the

PLC logic solver? Sometimes these can be lumped in with the sensors, but usually they are part of the logic solver.

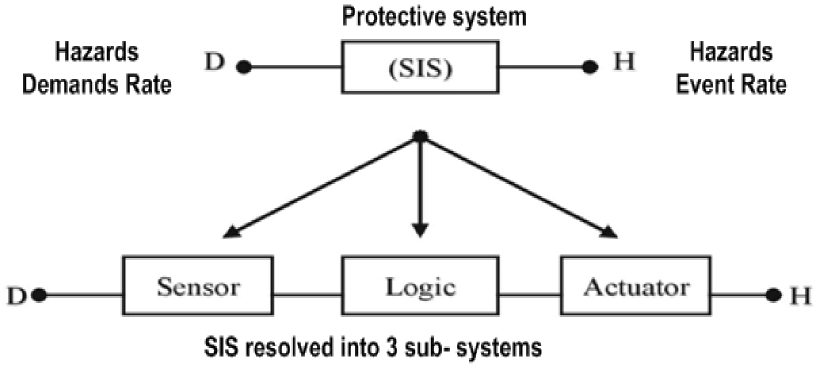


Figure 7.10
The SIS must first be resolved into 3 subsystems

Once defined, each subsystem must be evaluated individually for its ability to meet the SIL target. Because they are working in series, any failure of a sub-system will contribute to the failure of the overall SIS. Each subsystem must be qualified to meet the SIL and when joined together, the overall SIS must still satisfy the SIL target for PFD.

7.4. Selection of sensors and actuators for safety duties

7.4.1. Field devices for safety

Field devices basically comprise the sensors to provide the input information to the logic solver and the actuation devices required to carry out the trip function when the demand comes along. They are supported by the wiring and process connection arrangements. Taken together, they comprise the area with the greatest potential for problems. They are also the area where most engineers have a chance to exercise their skills and judgment in design, selection and maintenance.

7.4.2. Sensor types

Sensors are devices that can represent the value or status of a chosen parameter in a form suitable for decision-making in the SIS logic solver. Some illustrations are shown in Figure 7.11.

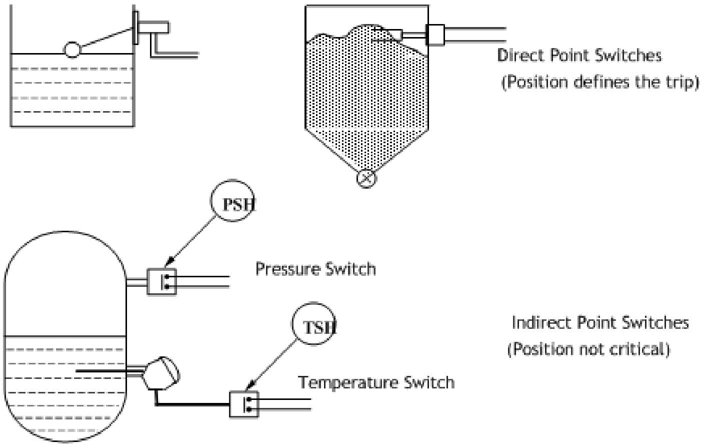


Figure 7.11
Categories of sensors

There are two basic categories of sensors: switches and transmitters that are discussed below.

A list of potential causes of failures in sensors: It is helpful to have a list of potential failures in sensor systems for reference in each application. Such a list also helps to identify ways in which the EUC control system can fail and create a hazardous event.

- Components of the instrument
- Accidental isolation
- Hostile process conditions
- Wiring errors
- Environmental electrical
- Calibration errors
- Response time
- Power supplies
- Lightning damage

7.4.3. Actuator types

The term actuator can be misleading. The devices used by the SIS to actuate the protective function should more correctly be called final elements.

Final elements are most commonly seen as process valves with air or hydraulic power to actuators that are spring loaded to close on release of the actuator fluid (typically described as "air to open, spring to close"). Or they are simply motor starter contactors that must be de-energized to break the power supply to the drives that must be tripped.

In large power control applications or in large valve applications, the final elements may have to use active power to carry out its trip task. These applications require backed up (i.e. redundant) power supplies to operate heavy-duty power isolating contactors or to drive motorized valves. In such cases the power system becomes an integral part of the final element and would require diagnostic monitoring.

In all cases there must obviously be a high degree of assurance that the final element will do the job when called upon to act. Hence there will be an emphasis on diagnostics and regular proof testing.

7.4.4. Guidelines for application of field devices

The application of any measurement and control device for duties in a safety instrumented system must take into account two primary considerations:

- The device must be applied using the best design techniques to minimize failures
- The device should meet the design requirements of IEC 61508 or IEC 61511

Given the diversity of failure modes, the design techniques required to minimize the fail-to-danger rate will be application dependent. The ground rules for design to minimize dangerous failure rates include the following techniques:

Fail-safe design: Design sensors and actuators to result in fail safe responses to their most likely failure modes. Then review spurious trip rates to see if they are acceptable.

Separation: Ensure separation between BPCS and SIS sensor/actuator systems as far as practicable.

Diagnostics: Search for ways of introducing diagnostics to frequently confirm the healthy operation of the device.

Redundancy: Use redundancy where a reduction in fail to danger rate is needed or where a low spurious trip rate is essential.

Diversity: Search for diversity of sensors where the risk of common cause failures is significant.

Armed with the knowledge of failure possibilities and knowing the redundancy rules we should now be ready to specify and select particular types of instruments for any given application. But, firstly we must remember that whatever instrument we use it must be qualified for use in our SIS.

7.4.5. Instrument selection

The IEC standards do not attempt to dictate what instruments to use. The ISA standard does offer some basic guidelines. The selection is almost entirely dependent on the application; hence it is not practicable to list any hard and fast rules about what instruments to use. The following points have been gleaned from the ISA standard and from personal experience.

Flow meters

- Vortex Shedding and Magnetic Flow meters are preferred due to proven performance
- Head-type flow measurements are to be avoided, if possible, due to impulse line problems – leakage, condensation, freezing and drift

Temperature sensors

- RTD and Thermocouple types: require burnout detection and alarm

- Be careful to locate sensors properly
- Ensure probes are seated in thermo wells
- Avoid thermistor types
- Infra red types with diagnostics

Pressure

- Gauge, differential pressure and absolute pressure types are very reliable
- Ensure range and trip setting are compatible
- Ensure impulse lines do not risk condensate build up or blocking
- Use remote diaphragm seals instead of long leg links, but: beware of drift; beware of vacuum effects on diaphragms

Level

- Wide range of devices – good possibilities for diversity
- Check process for effects of aeration, entrained solids, density shift
- Differential pressure types are reliable but prone to process effects
- Ultrasonic and radar types have smart electronics, hence:
- Check compliance with IEC rules for PES
- Exploit diagnostics
- Check process for foaming, vapor, boiling
- Nucleonic types – reliable, often with diagnostics

Analyzers: reliability and proven performance

- Gas detectors often used in redundant modes
- Lower for diagnostic facilities
- Use comparative process signals where possible to support diagnostics

Limit switches

- Use best quality industrial grade
- Sealed contacts where possible
- Proximity switches available with diagnostics
- Check leakage currents in OFF state to avoid false ON condition at SIS

Selection factors

- Material compatibility
- Shut off duty, shock loadings, leakage, fire resistance
- Speed of response
- Element types most used are: Ball, Gate, and Globe
- Butterfly valves used on air systems, large sizes
- Spring/actuator performance margins. Fail close/open
- Requirements for diagnostics
- Requirements for limit switch and/or position transmitter
- Sharing of valve duties with BPCS
- Cost

Solenoids: Critical components

- High-grade versions only, Stainless steel bodies

- Rated for outdoor service – sunshine to snow or enclosed
- Venting capacity to meet speed of response
- Use direct acting types – avoid pilot operated
- Use direct mounting, avoids risk of pinched vent lines

Installation design features

- Direct connections to process for field sensors, separate taps and impulse lines
- Dedicated cables, junction boxes, airlines and termination panels
- Dedicated power supplies
- Identification such as painting and labels
- Devices to have local indication to assist proof testing.

It is important for keeping the safety integrity of field devices in order that details of the installation do not lead to accidental malfunctions. It is useless if a perfectly good pressure transmitter can be accidentally shorted out or cross connected to another circuit. Hence segregation of shutdown system wiring and special identification are features that will repay the extra effort involved.

7.4.6. Technology issues

This section considers the impact of new technologies on SIS field device practices. The two items that have a major impact on field devices are instruments with self-testing diagnostics and bus communications. We need to consider how these technologies fit in with SIS concepts.

Intelligent field devices: Advantages and disadvantages

Intelligent instruments offer safety systems the advantages of being able to perform better quality measurements supported by internal diagnostics. Self-testing and safe responses to faults will help increase the safe failure fraction of a field device.

There are disadvantages for safety systems. Firstly, there are the general reservations about the risks of programmable systems in safety applications. These include:

- The potential for systematic errors in the software
- User configurations that may create new untested versions of the instrument
- The unauthorized in-service changes to settings, zero, range, mode etc.

One of the main purposes of IEC 61508 and IEC 61511 was to address these types of issues and find ways of dealing with them. Hence with the aid of safeguards based on IEC 61508 through design or IEC 61511 through prior use it becomes possible to use intelligent instruments in a safety system provided we stick to the rules. In brief, the answers to the above possible problems are:

- Instruments using PES should be manufactured using hardware and software engineering procedures in accordance with IEC 61508
- Limited ranges of software instructions should be made available to the end user to program the instrument within a tested range of configurations

- The program of the instruments should be password protected

Unless the above requirements are met, the fault tolerant rating of the instrument is loaded down by comparison with a non-PES version (see Redundancy section).

7.5. Selection of safety controllers

The logic solver stage of an SIS is where all the decisions are made to execute the safety function. Some SIS designs can actually operate without a logic solver if the sensors have the ability to directly operate the final elements. For example, a high-pressure trip switch connected directly to a solenoid valve releasing a shut off valve. However, in most applications the need for additional logic or signaling dictates that some form of logic function must be performed.

Traditionally, relay systems have been used for the logic function and these remain an attractive option for simple applications. In practice, most process plants, with hazardous materials, find the need for several safety functions and with this comes the need to manage the safety trips in an efficient manner. The need for interfacing to the operator increases and often there is a linkage between one trip function and another. As complexity grows, so does the need to have efficient supervision and control of the trip functions and their equipment. There is a need to be sure that all logic functions perform in such a way that start up and other plant operations interact smoothly with trip systems - the sophistication of the safety function begins to increase. These pressures, as well as the concern that relay systems are difficult to make secure and reliable as they grow in siz,e have caused many users to move first to solid-state hardwired logic and then to PLC based logic solvers.

The need for PLCs to have special protection against unpredictable failure modes and untested software logic combinations was realized many years ago and this has lead to a range of possible solutions being developed. The new standards addressed many of the potential problems of using basic PLCs for safety and have delivered a range of requirements that are usually satisfied by the specially designed safety PLCs. IEC 61511 has also addressed the issue of process plant equipment where a standard PLC has been provided with certain safeguards and has proved itself to be effective and reliable for safety. Rather than declare these PLC's unacceptable, it has made provision for these to be used with the safeguards being well defined.

With these choices available, the end user shall have a reasonable knowledge of the characteristics of the principal types of logic solver systems on the market in addition to the ability to relate these features to the needs of the project. This chapter outlines the operating principles of various types of logic solver, which should assist, with the initial steps of selecting the type and scope of logic solver to be used for the project.

7.6. System integration and application software

The objective of this module is to provide some guidance on how to deal with the application software stages of a SIS project. From the point of view of the end user of a safety PLC system, the software can appear to be deceptively simple. As soon as PLCs began to appear in safety systems, the specialists realized that there was a great danger that systematic design errors could be introduced into the logic

solver through problems with operating systems and through errors in the application logic.

The problem with software in safety systems is that it can be very difficult to control exactly what has gone into the system. Unacceptable in a safety system is an unpredictable response to a hazard demand.

IEC 61508 was the first major standard to lay down basic concepts and requirements for the control of systematic errors in software for safety applications. Part 3 provided comprehensive details of quality assurance procedures designed for the development of embedded operating system software as well as for the application layer. The impact of this approach is found in the availability, for end users, of certified operating systems supported by certified programming packages.

IEC 61508 part 3 is suitable for use by a manufacturer specializing in safety system products but is far too complex for the average instrument project team seeking to configure the logic for its safety functions. IEC 61511 has provided a more practical set of requirements and guidance notes designed for the application stage of a safety project while still following all the essential requirements of IEC 61508.

7.7. Programming tools

For most process plant applications, the engineer will require access to a configuration package supplied as part of the PES solution package. It may be helpful here to know the basic features to look for in a programming package.

It is a fundamental requirement of a safety certified PLC that the integrity of the whole software package is assured. Thus the embedded software and the application tools should be developed by the same vendor and should be proven to meet all the constraints of the IEC standard when operated as a complete facility.

IEC 61511 part 2 describes the programming and support tools as “the developers’ workbench”.

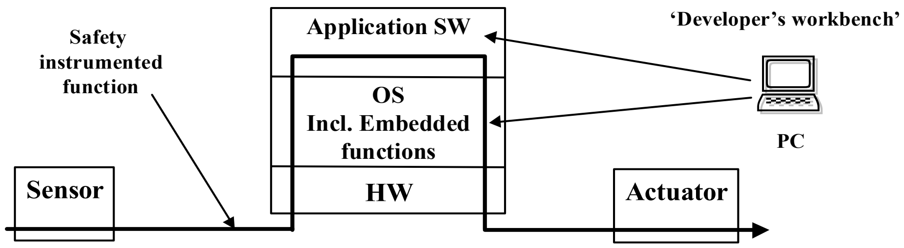


Figure 7.12
Programming tools

Tasks

- Configures logic solver I/O and communication subsystems
- Programs the logic and arithmetic functions of the application
- Facilitates testing of applications

7.8. Machinery safety

The safety of machinery affects all of us in everyday life, at home or at work or at leisure. Machines are part of our lives and our safety is dependent on the machines being safe for us to use at all times.

The following are the aspects that show how a machine should be made safe.

- **Physically Safe:** No sharp edges, spikes or projections that can be bumped into. No chance of it falling on someone. No ways in which it can throw objects around or let out jets of steam or noxious gases. No chance of explosions or radiation.
- **Mechanically safe:** The moving parts must not be able to hurt someone. If there's a risk that this can happen, then protection measures are required; fixed guards, movable guards, area sensing devices that stop the machine quickly if someone is in the danger zone.
- **Electrically safe**: There must be no chance of an electrical shock or a dangerous electrical circuit arrangement.
- **Functionally safe:** All the stops switches, guards and safety sensing devices are there to protect us must function properly. All safety controls that prevent movement at the wrong time must be reliable.

Safety measures based on sensors and control systems that are designed to ensure safe working of the machines are also known as Safety-Related Electrical Control Systems (SRECS).

7.8.1. Machinery and controls

Any assembly of devices designed to protect people from hazards or injuries that could arise from the use of the machine can be considered to be a **machinery safety system**. The machinery safety system may also provide protection for the machine itself or other machines against damage due to malfunctioning of the machine.

Fixed guards are usually the first line of defence. These prevent a person from being hurt by the machine, but in many cases the situation will require a logical action from the control system to prevent movement or other physical events from happening until safe conditions are proved to exist. These protective measures are the "safety functions" to be provided by the control system. Those parts of the basic control system, as well as any specially provided safety parts, are known as the **"safety related parts of the control system".**

Safety related controls include all parts involved in the safety function. In other words; the sensors, logic or evaluation units and the final drive interlocks and contactors or valves all belong to the safety control system.

While some safety devices can simply be passive guards such as shields or covers, it is most likely that many of the safety functions will be provided by a combination of mechanical devices and a safety related electrical control system. (SRECS).

7.8.2. Distinction between Machinery and Process Safety Control Systems

For process technology, the identification of unacceptable risks leads to a set of risk reduction measures that often include what is known as a safety-instrumented system. (SIS) or emergency shutdown system.

- Process plant shutdown systems define the grade or performance of their applications in terms of safety integrity levels or SILs.
- Machinery safety systems have been traditionally defined for performance by "safety categories" but will in future be moving to the same basis of SILs for complex and/or programmable safety systems.
- Process plant safety is subject to different regulations and design standards from those applicable to machinery safety, but the basic principles are essentially the same.

Some interesting questions arise when a section of process plant has a large and dangerous machine in the plant.

- Is the hazard coming from the process or from the machine?
- Which regulations are applicable?
- What design standard shall we apply?

If the hazard is due to the process, the plant safety systems can deal with it. If the machine presents hazards of its own the safety requirements will fall under machinery safety regulations.

7.9. Guide to Regulations and Standards

This topic provides outline information on the typical safety-relevant regulations and legal requirements applicable to both the supplier and the end users of machinery.

The legal aspects of moving machinery safety are generally arranged to cover two primary stages in the life of a machine.

- **Safe Manufacture –** this includes design and installation of plant and equipment.
- **Safe Operation –** this includes maintenance and modification of plant and equipment.

Generally, Laws and Regulations deal with these two aspects separately although regulators are increasingly aware of the need to improve the links from the supplier to the user.

Essentially, the safety of machinery must be tackled from two sides:

- The manufacturers and suppliers must build machines that are safe to use
- The users of machines (i.e. employers and workers) must ensure the machines are used in a safe manner and the workplace environment is safe for the workers.

Principles of EU Directives :

The sources of EU law through which the EU regulations are implemented can be divided into three categories,

- **Primary sources** - comprising the founding treaties, Community Acts (such as SEA) and further treaties (such as Maastricht or accession treaties).
- **Secondary sources** - comprising of regulations, directives and decisions. Through this EU implements the policy in more detail.
- **Non-legally-binding** - sources-opinions and other non-treaty acts (such as guidelines, resolutions, communications etc.)

Supply side laws: the EU "New Approach" Directives :

The term "New Approach Directives" applies to a range of EU directives that have certain basic principles in common. These include:

- Mandatory essential provisions, which apply to the product.
- Requirements for member states to ensure that products not in conformity with essential provisions are not allowed to circulate within the member states.
- The manufacturer is provided with the opportunity to certify conformity with the relevant directives. This leads to the manufacturer placing the CE mark on his product if it is claimed to conform.
- Legislation no longer specifies that specific standards have to be met. However, It can be "reasonably assumed" that when Harmonized standards are met, the associated goals of the EU directives are fulfilled.
- The manufacturer achieves conformity by compliance with a national law or regulation. This implies compliance with the equivalent laws in any other member state.

The CE mark is the characteristic symbol shown in the Figure 7.. This is applied to a product when the manufacturer or an appointed body certifies that the product conforms to the requirements of all applicable EU Directives.



Figure 7.13
The standard CE mark signifying a claim to conform to relevant directives

The products, which come under European Directives, and are to be placed on the market in the EU, must bear CE marking as it is a legal requirement. The affixing of CE marking to machinery by the manufacturer is to show its conformance to

that essential requirement as per 'New Approach' European Directives. The CE mark shall be distinct, visible, legible and indelible.

7.9.1. Some Machinery Safety Standards

7.9.1.1. Type A Standards – Basic Standards

These provide essential information for all machine builders. Generally there are three standards, which relate to machine safety:

- EN 414 - Safety of machinery: Rules for the drafting and presentation of safety standards. This defines the way standards are to be written.
- EN 292 Parts 1 & 2 - Safety of Machinery: Basic concepts, general principles for design. This defines the concepts of machine safety and specifies the general principles and techniques to help machine designers achieve safety. It incorporates in Annex 1 the EHSRs defined by the Machinery Directive
- EN 1050 - Safety of Machinery**:** Principles for risk assessment. This defines how to assess the risk of injury or damage to health, so that appropriate safety measures can be selected.

7.9.1.2. Type B Standards – Group standards

B standards are subdivided into two groups,

- Group B1: These cover higher-level safety aspects for design and are always applicable. E.g; ergonomic design principles, safety distances from potential sources of danger, minimum clearances to prevent crushing of body parts. Examples of these are EN 294 on safety distances and EN 563 on temperatures of touchable surfaces;
- Group B2: These cover safety components and devices for various machine types. These are applied when required. E.g; emergency stop equipment, two-hand controls, interlocking/ latching, non-contact protective devices, safety-related parts of controls. EN 281, on the design of pedals, is an example.

Standard EN 954 – 1 is of particular interest here, to control and electrical engineers, because it provides a method of defining the safety related parts of control systems used in machinery applications. This is the standard that defines our control system functions into safety categories such as 1,2, 3 or 4. It has close similarities to the safety categories used in safety instrument systems used for process control where they are known as Safety Integrity levels or SILs.

7.9.1.3. Type C Standards - Product Standards

These identify specific types or group of machines and involve the machinery-specific Standards. These inform machine manufacturers and users about the specific safety precautions they should take and safety devices they should use. e.g. for machine tools, woodworking machines, elevators, packaging machines, printing machines etc.

Chapter 8. Hazardous Areas and Intrinsic Safety

8.1. Introduction

A HAZARDOUS area / environment may be potentially prone to:

- An accident
- The creation of a dangerous situation
- Being beset with dangerous situations
- Flash points
- Being explosive
- Being inflammable or flammable
- Being radioactive etc

Hazards can be defined as the following:

Urban Structure Fires: Perhaps the most common human-caused hazard (often a disaster) is fire in large occupied buildings. Causes can be accidental or deliberate, but unless structures have been built to safe fire standards, and sound emergency procedures are used, heavy loss of life can result.

BLEVE: An entire community was involved at Mississauga, Ontario, Canada when 250,000 had to be evacuated to avert disaster following a train accident which triggered a series of BLEVE (Boiling Liquid Expanding Vapour Explosions).

Other Explosions: Great loss of life occurred in Halifax, Nova Scotia, Canada in 1917 when a ship carrying explosives collided with another. Australia's most disastrous explosion was in the Mt Kembla mine, Wollongong, in 1902, when 95 miners died. One of the worst non-mining explosions occurred in 1974 at the Mt St Candice Convent in Hobart, when seven died in a boiler explosion.

Toxic Emission: During 1984 cyanide gas escaped from a fertilizer factory in Bhopal, India. The resulting deadly cloud caused the deaths of approximately 2,000 people living close-by. In Australia in August 1991, the Coode Island fire burnt 8.6 million litres of chemicals in the heart of Melbourne and loomed as a potential disaster.

An area may also be considered "hazardous" for various other reasons. These may also include the use of electrical equipment in the vicinity of water, the risk of personal injury from moving or falling parts, or even the presence of biological hazards.

8.1.1. Basis of area classification

The first step in Area Classification is to list and identify the areas in the plant where there is the possibility of a conducive atmosphere for explosion or fire to occur. Based on the knowledge so acquired, the design, selection and operation of the equipment has to be influenced in such a way that the risk of fire or explosion taking place is minimized.

It is useful to understand what a Non-hazardous (safe) Area is.

“An area classified as non-hazardous has a small probability of a flammable mixture being present. It is also called a "safe area" and includes most control rooms.”

Area classification cannot be set rigidly into any standard. Each installation will be different in some respect and therefore each site must be examined on its individual merits.

There are three situations that can occur in an operating plant with reference to hazardous areas:

- A situation where an explosive atmosphere is present always or for long periods because of operational requirement, i.e. continuous.
- A situation where explosive atmosphere occurs frequently, or if infrequently may persist for a considerable time, i.e. primary.
- A situation in which explosive atmosphere occurs rarely and normally results from failure of equipment or procedures, i.e. secondary.

The above criteria are used in classifying the areas under ‘Source of Release’ methodology of classification.

Some typical examples of sources of release are:

- Open surface of liquid
- Virtually instantaneous evaporation of a liquid (for example from a jet or spray)
- Leakage of a gas mixture
- principles governing the sources of release
  - Sources giving a continuous grade of release
  - Sources giving a primary grade of release
  - Sources giving a secondary grade of release

8.2. Zonal Classification

8.2.1. Gases, vapor and mists

Areas where there is the likelihood of the presence of explosive gas-air mixtures are referred to as zones. Zones are classified as shown in the table below. The higher the number in this 'Zonal classification' the smaller is the risk of an explosion. This is as per IEC 79:

Zone 0	An area in which an explosive gas/air mixture is continually present or present for long periods
Zone 1	An area in which a gas/air mixture is likely to occur in normal operation
Zone 2	An area in which a gas/air mixture is not likely to occur in normal operation, and if it occurs, it will exist only for a short time.

8.2.2. Dusts

In respect of dust, the situation had been much more fluid. In recent times effort has been made to address this by classifying the Zones in a way which is similar to that adopted for gas and vapour.

Zone 20	An area in which combustible dust, as a cloud, is present continuously or frequently, during normal operation, in sufficient quantity to be capable of producing an explosive concentration of combustible dust in mixture with air, and / or where layers of dust of uncontrollable and excessive thickness can be formed.
Zone 21	Zone 21 is a Zone not classified as Zone 20 in which combustible dust, as a cloud, is present continuously or frequently, during normal operation, in sufficient quantity to be capable of producing an explosive concentration of combustible dust in mixture with air.
Zone 22	Zone 22 is a Zone not classified as Zone 21 in which combustible dust, as a cloud, is present continuously or frequently, during normal operation, in sufficient quantity to be capable of producing an explosive concentration of combustible dust in mixture with air.

Generally it is considered that 1 mm or less thickness of dust is not likely to result in the formation of an explosive atmosphere.

In carrying out an area classification, it is necessary to:

- Identify those parts of the plant where flammable dust can exist including, where appropriate, the interior of process equipment
- Assess the likelihood of occurrence of a flammable atmosphere thereby establishing the appropriate zonal classification
- Delineate the boundaries of the zones taking into account the effect of likely air movement
- Take into account, when assessing the area classification of a plant, the influence of the classification of adjacent plants

8.3. Area classification

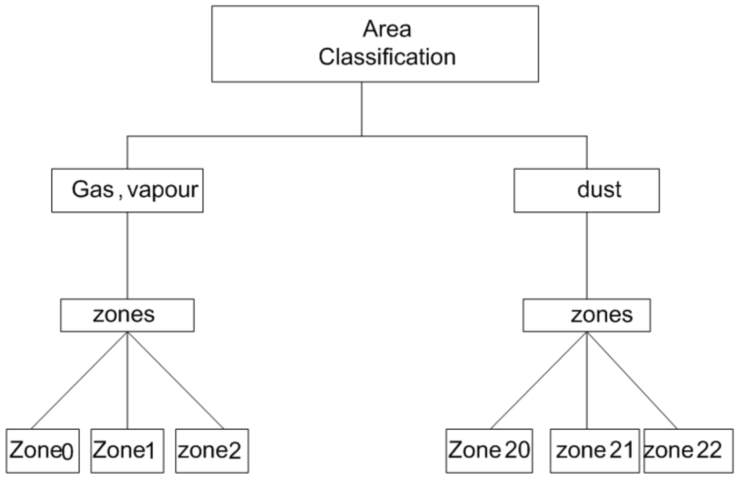


Figure 8.1
Area classification

8.3.1. Area classification – Gas and vapors

Area classification is a method of analysing and classifying the environment where explosive gas atmospheres may occur so as to facilitate the proper selection and installation of apparatus to be used safely in that environment, taking into account gas groups and temperature classes.

8.3.2. Sources of release

The basic elements for establishing the hazardous zone types are the identification of the source of release and the determination of the grade of release.

Since an explosive gas atmosphere can exist only if a flammable gas or vapour is present with air, it is necessary to decide if any of these flammable materials can exist in the area concerned.

If it is established that the item may release flammable material into the atmosphere, it is necessary, first of all, to determine the grade of release in accordance with the definitions, by establishing the likely frequency and duration of the release.

Having established the grade of the release, it is necessary to determine the release rate and other factors, which may influence the *type and extent* of the zone.

8.3.3. Type of zone

The likelihood of the presence of an explosive gas atmosphere and hence the type of zone depends mainly on the grade of release and the ventilation. A continuous grade of release normally leads to a zone 0, a primary grade to zone 1 and a secondary grade to zone 2.

8.3.4. Extent of zone

Consideration should always be given to the possibility that a gas which is heavier than air may flow into areas below ground level for example pits or depressions and that a gas which is lighter than air may be retained at high level, for example in a roof space.

The penetration of a significant quantity of flammable gas or vapour into the area can be prevented by suitable means such as:

- Physical barriers
- Maintaining a static overpressure in the area relative to the adjacent hazardous areas
- Purging the area with a significant flow of air.

Release rate of gas or vapor: The greater the release rates the larger the extent of the zone. The release rate itself depends on other parameters, namely:

- Geometry of the source of release
- Release velocity
- Concentration
- Volatility of a flammable liquid
- Flashpoints of flammable liquids
- Liquid temperature.

Lower explosive limit (LEL): For a given release volume, the lower the LEL, the greater will be the extent of the zone.

Ventilation: With increased ventilation, the extent of the zone will be reduced. Obstacles, which impede the ventilation, may increase the extent of the zone.

Relative density of the gas or vapor when it is released

- The horizontal extent of the zone at ground level will increase with increasing relative density and the vertical extent above the source will increase with decreasing relative density.

8.3.5. Openings

Openings between areas should be considered as possible sources of release. The grade of release will depend upon:

- The zone type of the adjoining area
- The frequency and duration of opening periods
- The effectiveness of seals or joints;

Openings are classified as A, B, C, and D with the following characteristics:

- Type A: Openings not conforming to the characteristics specified for types B, C or D
- Type B: Openings that are normally closed (for example automatic closing) and infrequently opened and which are close fitting
- Type C: Openings normally closed and infrequently opened, conforming to type B, which are also fitted with sealing devices
- Type D: Type D openings are effectively sealed, such as in utility passages (for example ducts, pipes) or can be a combination of one

opening type C adjacent to a hazardous area and one opening type B in series.

8.3.6. Ventilation

Ventilation, i.e. air movement leading to replacement of the atmosphere in a (hypothetical) volume around the source of release by fresh air will promote dispersion. Two main types of ventilation are thus recognized**:**

- Natural ventilation
- Artificial ventilation, general or local.

8.4. Methods of explosion protection

When electrical equipment is to be located in a hazardous area it must be designed manufactured and certified for that purpose. There are several methods of protection available and these are based upon the various protection techniques.

The zonal classification of the hazardous area that the equipment is to be located in will, or partially, determine the equipment's method of protection. This electrical equipment is known as 'explosion protected' equipment, the symbol being Ex, or as per CENELEC standards EEx, followed by the letter designating the mode of protection. Care must be taken not to confuse the term 'explosion protected' with the North American term of 'explosion proof' used to describe their hazardous area equipment. Each technique of protection is assigned a code letter depicting the type of protection.

Each technique of protection is assigned a code letter depicting the type of protection.

8.4.1. Exclusion of the explosive atmosphere (criterion a)

This is when the gas / air or vapour / air mixture is prevented into coming into contact with components or equipment that could cause ignition.

Pressurized (Ex 'p') - if clean dry air or an inert gas is pumped into an enclosure housing electrical equipment and a positive pressure is maintained at 50Pa with respect to the surrounding atmosphere then flammable gas or vapour will be excluded.

Purged (Ex 'pl') - similar to Ex 'p' except that an air flow or inert gas flow is maintained in an enclosure to ensure that there is no build up or presence of a flammable gas or vapour.

Ventilated (Ex 'v') - used in large areas to dilute flammable gas or vapour to well below L.E.L. and to reduce the temperature of electrical equipment by airflow passing over the equipment.

Encapsulation (Ex 'm') - the main requirement for encapsulation is that the apparatus to be protected is encapsulated in resin with at least 3 mm of resin between it and the surface.

8.4.2. Prevention of sparking (criterion b)

This involves selecting components or equipment that will not provide a source of ignition when in normal use.

Increased safety (Ex 'e') - perhaps the most widely used method of protection.The design and manufacture of this equipment assures safety against ignition through ensuring that the temperature of the equipment will not become excessive and that the incidence of arcs and sparks in normal service is prevented.

Non-sparking (Ex 'n') - apparatus which does not produce arcs, sparks or hot surfaces in normal operation are considered within this scope and this system of protection is common amongst three-phase induction motors used in hazardous areas.

8.4.3. Explosion containment (criterion c)

If a gas / air or vapour / air mixture manages to enter an enclosure that contains electrical equipment and that mixture is ignited then this enclosure must be robust enough to contain the explosion and ensure that the escaping products of the explosion do not cause ignition outside of the enclosure.

Examples:-Flameproof (Ex 'd') , Sand filled (Ex 'q') .

8.4.4. Energy limitation (criterion d)

This involves the limitation of energy into a hazardous area so that there is insufficient energy allowed into the circuit to cause ignition.

Intrinsically safe (Ex 'ia' and Ex 'ib') - this is a common type of protection where the required limitations of voltage and current allow its use. With intrinsic safety, there is always the need for a certified interface unit such as a Zener barrier or galvanic coupler to couple the supply from the safe area to the intrinsically safe equipment in the hazardous area.

8.5. Flameproof concept Ex d

This method is also widely used along with the equally popular Ex 'e' concept.

The basis of Ex 'd' equipment is that ordinary non-certified electrical apparatus such as relays, switches, terminal blocks etc are located in an enclosure. If a gas / air or vapor / air mixture enters the enclosure in sufficient quantities and ignites, the enclosure will contain the effects of the ignition.

This method is generally suitable for Zone 1 and 2.

8.5.1. Flameproof enclosure

The term FLP (mnemonic for flameproof) originates from the mining industry where it was first used.

A flameproof enclosure is defined in the standards as: ‘An enclosure for electrical apparatus that will withstand an internal explosion of the flammable gas or vapor which may enter it without suffering damage and without communicating the

internal flammation to the external flammable gas or vapor for which it is designed, through any joints or structural openings in the enclosure'.

A flameproof enclosure is designed to withstand the pressure of an internal explosion; it is not necessary therefore to provide openings for pressure-relief. Where there is a joint, however, or where a spindle or shaft passes through the enclosure, the products of the explosion can escape. It should be understood that the aim of a flameproof enclosure is not necessarily the total avoidance of any gaps in an enclosure. The misconception that it should be 'gas-tight' is misplaced. The principle recognizes that some openings are unavoidable in practice and so restricts itself to requiring that the size of such openings should not exceed the safe limit above which the nature of the escaping flame is such as to ignite a specified flammable atmosphere. On the other hand, it is not the aim to require joints to be deliberately spaced to give an opening.

Flameproof Joint: The place where the corresponding surfaces of the different parts of a flameproof enclosure come together; where the flame or products of combustion may be transmitted from the inside to the outside of the enclosure.

Length of Flame Path: The shortest path traversed by a flame through a joint from the inside to the outside of an enclosure.

Gap: The distance between the corresponding surfaces of a flameproof joint after the electrical equipment has been assembled.

Pressure Piling: A condition of rise in pressure resulting from ignition of pre-compressed gases in compartments or subdivisions other than those in which ignition was initiated and which may lead to a higher maximum pressure than would otherwise be expected.

Maximum experimental safe gap (MESG): Any path, which the flame or hot gases may take, needs to be of sufficient length and constriction to cool the products of the explosion so as to prevent ignition of a flammable atmosphere external to the enclosure.

8.6. Intrinsic safety

The Fire Triangle (Figure 8.2) analogy can be used to explain the objective of Intrinsic Safety. In the presence of the "most easily ignitable mixture" of a given flammable vapour with air, ignition cannot occur if the levels of heat or sparks are insufficient.

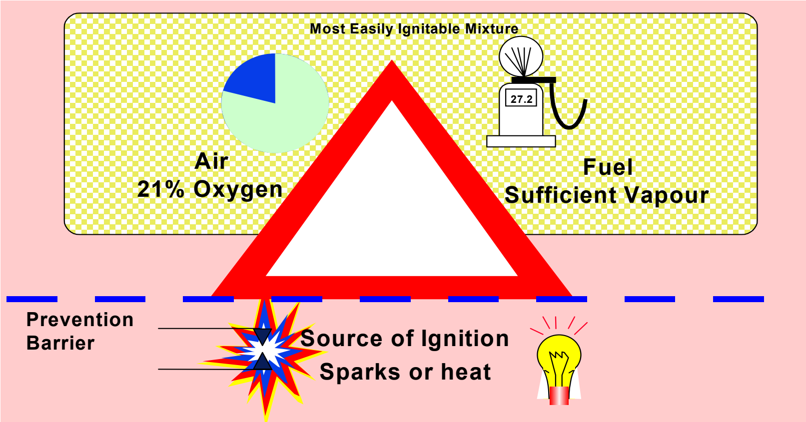


Figure 8.2
The fire triangle

The principle of IS to ensure that levels of heat or size of sparks that occur in an electrical circuit which comes into contact with a flammable gas, are limited to below those which will cause ignition.

An IS circuit is defined in Standard IEC 60079-11 as:

"A circuit in which any spark or thermal effect produced in the condition specified in this international standard, which include normal operation and specified fault conditions is not capable of causing ignition in a given explosive gas atmosphere."

The standard repeats and qualifies this where it declares that three basic criteria must be satisfied;

- Separation from other circuits
- Temperature classification; and
- The inability to cause ignition by sparking.

The definition suggests that sparks and heat are permitted in a circuit under specified fault conditions, but these must never exceed levels that could be incentive. These criteria and the requirement for separation from other circuits provide the high integrity necessary.

The use of the term 'circuit' has important implications. Electrical energy can only produce heat or sparks when electricity is flowing. Since it can only flow in a complete circuit, it is the safety of the circuit that is of concern. The components of the circuit do not pose a threat unless electricity is passing through them, in which case they must be part of a 'circuit'. Heat and sparks occurring in this circuit can be assessed for compliance with the standards for IS.

A 'circuit' can mean any of the following arrangements, increasing in complexity:

- A single cable looped through a hazardous area
- An assembly of electrical components working together as an electronic device (such as an instrument)
- A number of assemblies can be interconnected in the same circuit.

The circuit may operate with energy levels that are quite safe under normal conditions. Under probable fault conditions acting within or onto the circuit, the

circuit must still not be able to emit heat or sparks in sufficient quantities to cause ignition where it encounters a hazardous area. Internal faults and certain external faults must be adequately protected against.

The possible 'faults' are predicted by careful examination of what failure mechanisms could occur. Components are built into the design of the complete circuit in order to maintain energy to known safe levels under these fault conditions. These components are termed, 'safety components' and are in-built into the operation of the circuit. To further enhance the integrity, the failures of the 'in-built' components are separately assessed to ensure that if they fail in a specified manner, safety is still maintained.

8.7. Increased safety

The protection concept of increased safety is one intended for use in Zone I and less hazardous areas. It is generally denoted by adding suffix 'e', i.e., Ex 'e'.

Increased safety is a type of protection applied to electrical equipment that does not produce arcs or sparks in normal service and under specified abnormal conditions, in which additional measures are applied so as to give increased security against the possibility of excessive temperature and of the occurrence of arcs and sparks.

Electrical apparatus with type of protection increased safety 'e' is distinguished by the fact that it does not generate any ignitable sparks during normal operation. The aim of this type of protection is to avoid the occurrence of ignitable sparks and thus have a distinctly higher degree of safety compared with conventional electrical apparatus.

The Ex 'e' standard specifically does not permit the inclusion of any discontinuous contact. No switches or switching mechanisms are allowed in this concept of protection. Sparks therefore cannot occur and spark energy does not need to be considered.

This protection aims are mainly reached by applying the following principles:

- The enclosures are designed in such a way that the entry of moisture and dirt in hazardous quantities is prevented. The IP protection class IP 54 is laid down as the minimum requirement and the enclosures have a mechanical strength that can withstand the typical harsh operating conditions in an industrial plant. Enclosures must guarantee the minimum Protection standard IP 54 even under severe external mechanical forces
- Internally, the clearance and creepage distances must also be so dimensioned that even under harsh ambient conditions, no short circuits via creepage paths or flashovers can occur
- The electrical connection terminals are designed in such a way that it is not possible for the cable connected to them to come loose
- The dimensioning of the apparatus in electrical terms ensures that no inadmissible temperatures can occur inside or on outer parts of the apparatus.

8.8. Certification (components)

Component parts to be included in larger arrangements may be ‘component certified’ for some flexibility. In an Ex ‘e’ junction box for example, the enclosure will be impact tested. The terminals to be used within will be component approved.

The main uses of this technique are found in higher power circuits such as induction motors, fluorescent lighting fittings, junction boxes, and terminal housings. The German standards from which this came promote the use of toughened plastic cable sheaths on permanent installations as opposed to the more expensive steel wire armoured cable used elsewhere.

When applied to junction boxes, an Ex ‘e’ enclosure is given an ‘enclosure factor’ when certified. This represents the highest number of ‘terminal-amps’ permitted in the box. Terminals mounted in the box must be component approved. The total of terminal-amps must be calculated and must be equal to or less than the enclosure factor.

Construction requirements: The standards permit the construction of apparatus in such a way that during normal operation of the equipment, it is unlikely to become a source of ignition. The rules therefore seek to develop an acceptable level of integrity by considering standard industrial grade equipment and enhancing some aspects of its construction. This is as opposed to the inclusion of specific electrical or mechanical techniques to prevent ignition, which are applied in some of the other methods.

8.9. Principles of testing

As we have seen that increased safety apparatus calls for a high degree of integrity of material and manufacturing and hence its assessment involves checking that the manufacturer has complied with the design parameters of the standard.

Thus it follows that a testing program for any type of device should include the following as a minimum:

Test of creepage distances: This will involve determining the comparative tracking index of the insulating material. This is required to be done to establish the minimum distance, also known as creepage distance, required between the live parts or live parts and ground.

Special apparatus consisting of application of test voltages and ammonium chloride are applied to the material. Two electrodes spaced 4 mm apart are used to apply the test voltage. This determines the grade of insulating material and hence the resultant properties.

Ceramic material is exempted from this test.

Temperature-rise test: This test is carried out so as to determine the temperature class of the apparatus. Unlike flameproof apparatus, all surfaces, including internal surfaces of the apparatus, are considered. Exceptions would be internal components using exclusion or containment techniques, e.g. encapsulation or flame proofing.

For the apparatus to pass the tests not only for itself, but also for the components that are housed in it, they should also be within the limiting temperature range as per limits of the insulating material used. These limits are to be in line with industry Standards, e.g. cable insulation, or are given in AS 2380.6, e.g. insulation of motor windings.

Degree of protection tests: As this is an important aspect of protection and the apparatus has to meet various levels of protection against the ingress of solid objects or water are specified which must be tested in accordance with codes and standards as specified for degrees of protection provided by enclosures for electrical apparatus (IP Code).

8.10. Non Sparking concept

The most widely used forms of explosion protection, which utilize the technique of energy limitation, are non-sparking and intrinsic safety. While both share a common foundation, they do differ greatly in many aspects. These differences deal mainly with the application of safety factors. It is for this reason that non-sparking is limited to Zone 2 hazardous locations while intrinsic safety is acceptable for Zone 0, 1 & 2 locations. Each does have its merits, which is why it is not uncommon to see the two techniques used together.

The perspective of the user is that Ex ‘n’ is a less costly approach than IS because no interface (e.g. barrier or isolator) is required. It could be argued that the overall installation is less safe with Ex ‘n’ than with Ex ‘i’ with only a marginal cost saving. The use of Ex ‘n’ remains restricted to Zone 2 only. This raises the concern that area classification may be influenced in order to accommodate Ex ‘n’ apparatus.

8.10.1. Definitions

Type of protection ‘n’ is defined in the standard as:

“A type of protection applied to electrical apparatus’ such that, in normal operation, it is not capable of igniting a surrounding explosive atmosphere and a fault capable of causing ignition is not likely to occur.”

The general requirements of such apparatus are that it shall not, in normal operation:

- Produce an arc or spark unless
- The operational arc or spark occurs in an enclosed break device
- The operational arc or spark has insufficient energy to cause ignition of a flammable atmosphere
- The operational arc or spark occurs in a hermetically sealed device.

(It is to be noted that sliding contacts are considered to be sparking in normal operation.)

- Develop a surface temperature or hot spot capable of causing ignition of an external flammable atmosphere.

(It is to be noted that this requirement applies to the temperature of internal and external surfaces to which a surrounding atmosphere has access except internal

surfaces within enclosed-break devices, hermetically sealed devices or restricted-breathing enclosure.)

8.10.2. Principles of design

The 'non-sparking' concept of protection was originally accepted as safe on the basis that manufacturers used good quality and well designed industrial equipment with little or no additional requirements. This is if it is operated well within its rating and installed in areas where the risk of contact with a potentially flammable atmosphere was adequately low (Zone 2 only).

The Standard BS 4683: Part 3: 1983 in UK, AS 2380.9 in Australia and IEC 60079-15 internationally, eventually emerged to formalize the concept of good design using industrially graded apparatus and this was termed Type 'n' equipment. The standard clarified the method in that it permitted equipment to produce sparks or for the surface temperature of electrical assemblies to rise in temperature, but not to levels that could cause ignition.

More recently, the method was re-designated Type 'n' under BS 6941:1988. It has been updated and made more flexible to include the specific needs of instrumentation where circuits with less than 75V DC were recognized. Discontinuous contacts were permitted provided the resultant spark could be shown non-sparking. The same curves are now required to assess the safety of type 'n' circuits. The system is deemed safe in normal operation only and faults in the equipment or components and their effect on the explosion protection integrity are not considered. There are some parallels with Ex 'i' in that this technique is 'safe with no faults' and therefore could be likened to an unofficial grade of safety Ex 'ic' if the same logic applied to the IS technique is followed.

8.11. Concept Ex p

This is one of most popular and widely used protection concepts using 'the principle of separation' of three elements of the fire triangle in order to prevent ignition (see Figure 8.3).



Figure 8.3
Explosion / Fire triangle

This is a useful technique because an artificial 'safe area' having sufficient integrity can surround virtually any electrical equipment. The technique is flexible in that it can be adopted for many situations. The system is not used often due to:

- High cost, and
- Inconvenience of equipment accessibility of this solution

The system is expensive to operate and maintain because the clean air must be pumped and controlled by other equipment exposed to the hazardous area. The major benefit is that it can be used on very small enclosures up to complete control

rooms. The problems associated with this tend to complicate area classification. A guaranteed gas free air supply must be maintained so it must be piped in from a safe area. Disposal of used air, if it is likely to contain gas during the initial purge, must be handled such that it does not convert a safe area to a hazardous area.

Of the four techniques using the principle of separation, the Ex p technique is very versatile. Any uncertified equipment may be placed in an enclosure where an inert gas or air is supplied and maintained at a slightly higher than atmospheric pressure level. Ordinary air is most commonly used although there may be cases where nitrogen is preferred.

8.11.1. Definitions

Control of the atmosphere within a room or apparatus enclosure permits the safe use of electrical apparatus, which in the absence of the control would be unsuitable. The pressurizing or purging of the room or enclosure can achieve this. In some cases the two methods cannot be regarded as independent but for the purpose of this discourse the following definitions apply:

Pressurizing: It is a method of safeguarding, whereby air or inert gas in a room or enclosure, is maintained at a pressure sufficient to prevent the ingress of the surrounding atmosphere, which might be flammable. Where appropriate, the pressure may be provided by a mechanical ventilation system.

A variant of this is also known as static pressurization. With static pressurization, the overpressure is created before the system is commissioned by charging the enclosure with protective gas and maintained solely by the sealing of the enclosure without any protective gas being supplied in the hazardous area. The protective gas must be inert. A maximum oxygen concentration of 1 per cent by volume is permitted. Measuring equipment should be used to check this on every charging process.

Purging: This is a method of safeguarding whereby a flow of air or inert gas is maintained through a room or enclosure in sufficient quantity to reduce or prevent any hazard, which could arise in the absence of the purge. (To 'reduce' in this context means to reduce the risk of a flammable atmosphere occurring, thus permitting the use of electrical apparatus with a lower standard of safeguarding. Where the object is to 'prevent' a hazard, 'sufficient' shall take account of the highest likely rate of release of flammable material within or into the room or enclosure).

Where appropriate the purging may be provided by mechanical ventilation of the forced or induced type.

Pressurizing/ Purging: This is a method of safeguarding employing both pressurizing and purging.

Pressurization with leakage compensation is characterized by the fact that the required overpressure is established in the interior of the enclosure after purging. With closed outlet, a supply of protective gas (instrument air or inert gas) is sufficient to compensate for the leakage flow from the pressurized enclosure and pipelines.

In the case of pressurization with continuous flow of protective gas, the overpressure is achieved by the continuous flow of protective gas within the pressurized enclosure. Pressurization with leakage compensation and pressurization with continuous flow are based on a similar technical principle. However, the required protective gas flow rates differ greatly, leading to different designs. Due to its drawbacks in operation, static pressurization is not widely used. Continuous flow only offers advantages in the relatively rare case of internal release (analyzers).

8.11.2. Principles of application

The type of protection, pressurized apparatus "p" encloses electrical equipment or systems representing potential ignition sources in a tight enclosure. Instrument air or inert gas is introduced into this enclosure until a defined overpressure in relation to the external atmosphere is achieved, which is then maintained during the operation of the system. This overpressure prevents penetration of flammable gas or combustible dust from outside into the enclosure and hence the coincidence of an explosive atmosphere and an ignition source. In principle, pressurization technology is also used for online analyzers, which in turn may be used to analyze flammable gases (or liquids). In such cases, flammable gases are fed via a pipeline to the analyzer in a pressurized enclosure. Any leakage in these pipelines or even in the analyzer may constitute an internal source of flammable gases inside the pressurized enclosure.

The area containing the flammable gas (i.e. the pipelines and analyzer) is described as a containment system. Depending on the technical design of this containment system and gas feed system; it is described as an infallible containment system (no release), a limited release system with a predictable maximum release rate or an unlimited release system. With unlimited release, overpressure must be created by an inert protective gas, which prevents oxygen from penetrating the enclosure. With limited release, a sufficiently large volume of air is used to dilute the combustible gas outside a small "dilution area" so that an explosive atmosphere is unable to form.

Relative to the explosion triangle mentioned at the start, this means that, there is no explosive mixture inside the enclosure (there is no flammable gas, or it is only present in amounts below the lower explosive limit (LEL), or there is no oxygen). In principle, any non-explosion protected apparatus and systems may be installed in the enclosure.

8.12. Other protection concepts

There are additional concepts that are not so widely used, but have been developed and find their use in specific applications. These, when used in conjunction with one or more of the popular methods of protection, can really lead to economically safe solutions.

8.12.1. Ex 'o': oil filling

This is one of the separation methods where the oil is used as a separation medium. The Ex 'o' method shown in Figure 8.4 was originally conceived for

high power equipment. It provides explosion protection on a similar basis to Ex 'p'.

The definition of Ex 'o' protection concept in standards is,

'A type of protection in which the electrical apparatus or parts of the electrical apparatus are immersed in a protective liquid in such a way that an explosive atmosphere which maybe above the liquid or outside the enclosure cannot be ignited.'

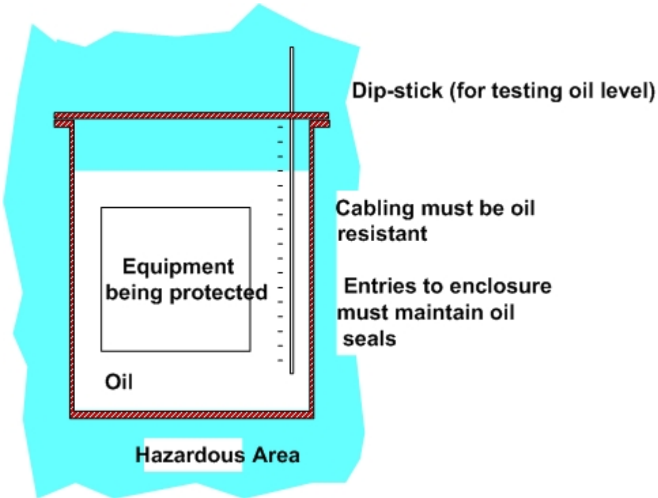


Figure 8.4
Ex o oil immersion

8.12.2. Ex 'q': Quartz/sand filling

This is another of the methods of protection under concept of separation.

'A type of protection in which the parts capable of igniting an explosive atmosphere are fixed in position and completely surrounded by filling material to prevent the ignition of an external explosive atmosphere.'

It is rarely used on its own and is mainly found in combination with the other construction techniques described here. It cannot be used in situations where movement is required, i.e., for the protection of relay contacts.

8.12.3. Ex 'm': Encapsulation

Encapsulation is used to prevent flammable gases from reaching a potential source of ignition in its own right. It may be thought of as the method of electrical protection using solids whereas Ex 'o' uses liquid and Ex 'p' uses gas.

The standards define it as follows:

'Protection of electrical components by enclosure in a resin in such a way that an explosive atmosphere cannot be ignited during operation by either sparking or overheating which may occur within the encapsulation.'

In other words encapsulation is used to prevent flammable gases from reaching a potential source of ignition within the encapsulated apparatus.

8.13. Earthing and Bonding

Correct "Earthing" is primarily required for the assurance of general electrical safety, reducing the risks to both human life and installations.

Electrical "earthing" is required for five main purposes:

- To reduce the risk of personnel shock
- To operate electrical protective devices
- To guard against lightning surges
- To control electrostatic discharge
- To minimize electrical interference.

8.13.1. Personnel safety

The effects of electricity on humans depend on the level of current and where it enters and leaves the body. Research shows that the limbs have a resistance of about 500 Ohms. The central torso has a very low resistance value owing to the high water content. The effect of electricity penetrating the skin can be liked to the characteristics of a zener diode with a reverse breakdown voltage of 5 to 10 V. This depends on the individual's skin characteristics and the tendency to dry or greasy skin.

8.13.2. Hazardous area considerations

Structural or fault currents arising from electrical equipment operating in hazardous areas must not become a source of heat or sparks. Equipment must be adequately earthed to ensure that connections are of high integrity and low impedance.

8.13.3. Definitions

There is a subtle but essential difference between earthing and bonding, which must be understood.

"Earthing" is where a low impedance path is provided in order for return currents to operate electrical protection devices such as fuses and over-current trips in an appropriately short time.

"Bonding" is where voltage differences between electrical conducting parts are eliminated.

International electrical supply regulations (ESRs) that cover fixed electrical equipment and installations require that there be an earth return that is backed up by a physical connection to "terrestrial" earth. In this way there are two return paths acting in parallel, which enhances the integrity of an earthing system. One path is an earth path because its primary function is to conduct fault currents. The other connection is to ensure that significant voltage differences do not appear between devices. Refer to Figure 8.5.

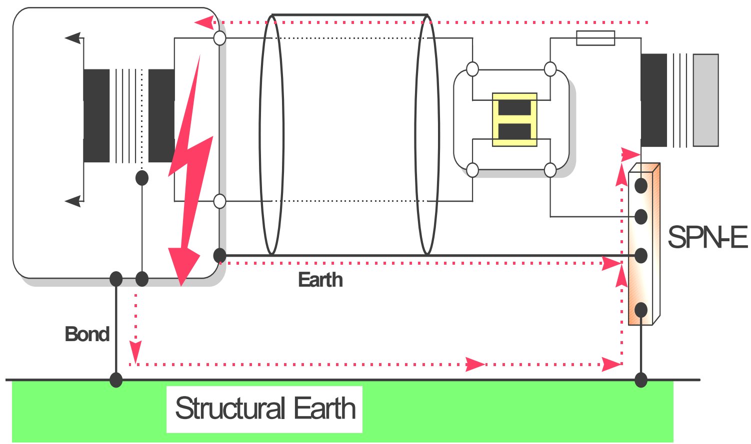


Figure 8.5
Earthing and bonding in parallel

The “Earth” and “Bond” conductors act in parallel. This is an advantage in that the two paths reduce the impedance. One path may be viewed as backing up the other lest it should fail.

8.13.4. Clean and dirty earthing

In any electrical system using ac supplies, current will flow in earthing and bonding paths. These are unavoidable and have to be coped with in the course of devising strategies. There are two reasons for this.

- Parasitic capacitance
- Fault currents

8.14. Standards and codes of practice

The Standards and Codes of Practice that apply in various countries are not always specific in their requirements for earthing as applied to IS and related topics.

In the UK, BS5345 states specific reference to the Star-point Neutral Earthing system of the incoming supply. In Germany, VDE 0165 requires the creation of a reference potential to which plant and supply are connected. A Canadian standard allow the designation of an Earth reference but does not state to what else it is to be connected to other than the Barrier earth.

8.15. Fault finding and repairs

There is no right or wrong way to fault find on instrument loops and systems. There are however, safe and potentially unsafe ways that are of the greatest importance to consider.

8.15.1. Fault finding routine

It is usual for a fault to be investigated by following a logical routine, an example of which is shown in Figure 8.6.

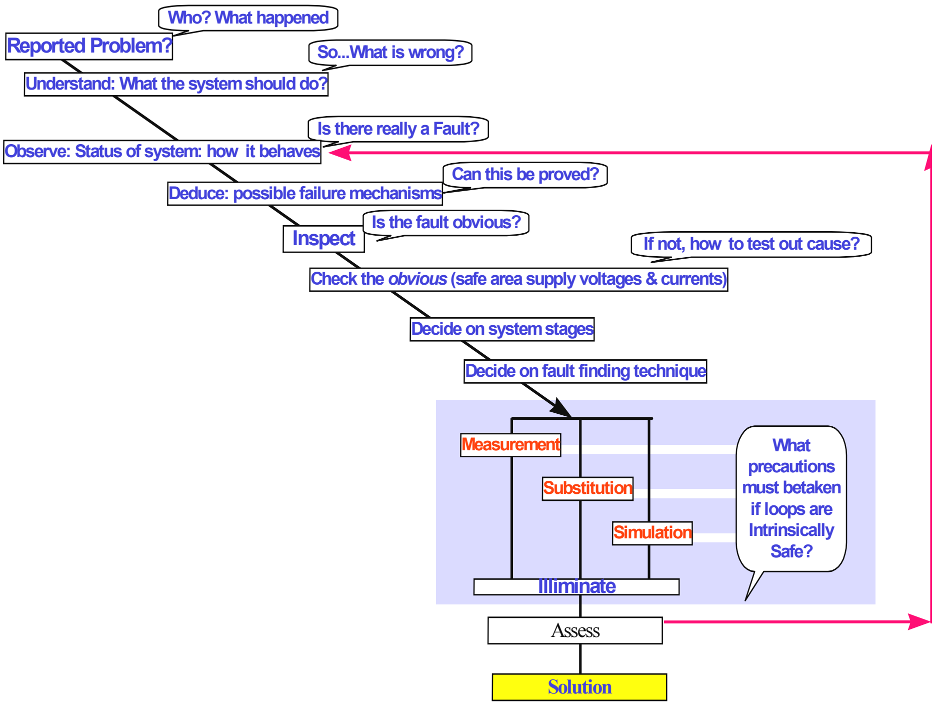


Figure 8.6
Possible fault finding routine

8.15.2. Safety assessment of testing

There is a legal 'duty of care' on all industrial personnel that adequate precautions are taken so as not to endanger life or investment during the course of work. Some assessment of work on electrical apparatus is required in order to ensure that the integrity of protection is not compromised.

8.15.3. Test equipment

There is a great variety of test equipment available on the market. It may seem obvious to point out that test equipment for hazardous area use will be either certified or uncertified.

The following are the types of tests conducted.

Insulation testing: Insulation testing is a requirement of all Standards. The test used is for the circuit to withstand 500Vac rms for one minute

Low voltage insulation test: The principle of this test is applying a low voltage source and monitoring the current taken with an adequately sensitive measuring instrument

The 500Vac test: The high voltage test to instrumentation should be performed as shown in Figure 8.7. This is preferably carried out in the safe area or if necessary in the hazardous area.

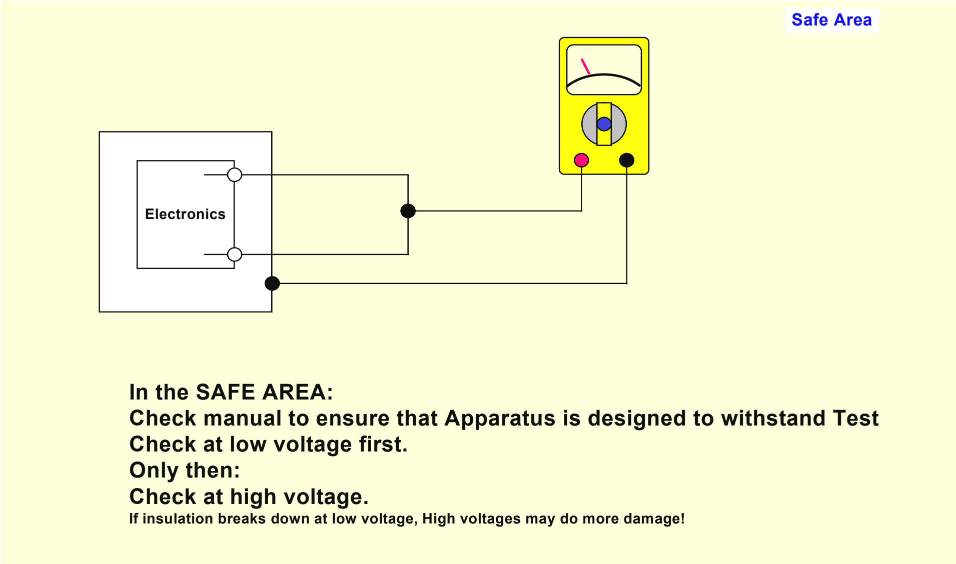


Figure 8.7
High voltage testing of instruments

Chapter 9. SCADA

9.1. Introduction and Brief History of SCADA

SCADA (Supervisory Control and Data Acquisition) has been around as long as there have been control systems. The first "SCADA" systems utilized data acquisition by means of panels of meters, lights and strip chart recorders. Supervisory control was exercised by the operator manually operating various control knobs. These devices were and still are used to do supervisory control and data acquisition on plants, factories and power generating facilities.

9.1.1. Fundamental Principles of Modern SCADA Systems

SCADA refers to the combination of telemetry and data acquisition. SCADA encompasses the collecting of the information, transferring it back to the central site, carrying out any necessary analysis and control and then displaying that information on a number of operator screens or displays. The required control actions are then conveyed back to the process.

The PLC or Programmable Logic Controller is still one of the most widely used control systems in industry. As needs grew to monitor and control more devices in the plant, the PLCs were distributed and the systems became more intelligent and smaller in size. PLCs and DCS or (Distributed Control Systems) are used as shown below.

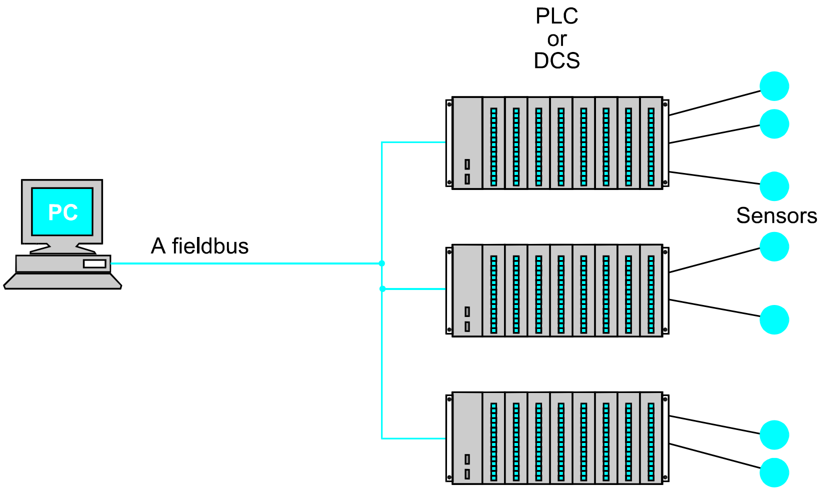


Figure 9.1
PC to PLC or DCS with a fieldbus and sensors

The advantages of the PLC / DCS SCADA system are:

- The computer can record and store a very large amount of data.
- The data can be displayed in any way the user requires.
- Thousands of sensors over a wide area can be connected to the system.
- The operator can incorporate real data simulations into the system.
- Many types of data can be collected from the RTUs.
- The data can be viewed from anywhere, not just on site.

The disadvantages are:

- The system is more complicated than the sensor to panel type.
- Different operating skills are required, such as system analysts and programmer.
- With thousands of sensors there is still a lot of wire to deal with.
- The operator can see only as far as the PLC.

9.1.2. SCADA Hardware

A SCADA System consists of a number of Remote Terminal Units (or RTUs) collecting field data and sending that data back to a master station via a communications system. The master station displays the acquired data and also allows the operator to perform remote control tasks.

On a more complex SCADA system there are essentially five levels or hierarchies:

- Field level instrumentation and control devices
- Marshalling terminals and RTUs
- Communications system
- The master station(s)
- The commercial data processing department computer system

9.1.3. SCADA Software

SCADA Software can be divided into two types, Proprietary or Open. Companies develop proprietary software to communicate to their hardware. These systems are sold as "turn key" solutions. Open software systems have gained popularity because of the Interoperability they bring to the system.

Citect and WonderWare are just two of the open software packages available on the market for SCADA systems

9.1.4. SCADA and Local Area Networks

To enable all the nodes on the SCADA network to share information, they must be connected by some transmission medium. The method of connection is known as the network topology.

Nodes need to share this transmission medium in such a way as to allow all nodes access to the medium without disrupting an established sender.

Ethernet is the most widely used LAN today because it is cheap and easy to use. Connection of the SCADA network to the LAN allows anyone within the company, with the right software and permission, to access the system. Since the data is held in a database the user can be limited to reading the information.

9.1.5. Modem Use in SCADA Systems



Figure 9.2
PC to RTU Using a Modem

Often in SCADA systems the RTU (Remote Terminal Unit (PLC, DCS or IED)) is located at a remote location. This distance can vary from tens of meters to thousands of Kilometers. One of the most cost-effective ways of communicating with the RTU over long distances can be by dialup telephone connection. With this system the devices needed are a PC, two dialup modems and the RTU (assuming that the RTU has a built in COM port). The modems are put in the auto-answer mode and the RTU can dial into the PC or the PC can dial the RTU.

9.1.6. System Implementation

When first planning and designing a SCADA system, consideration should be given to integrating new SCADA systems into existing communication networks in order to avoid the substantial cost of setting up new infrastructure and communications facilities. This may be carried out through existing LANs, private telephone systems or existing radio systems used for mobile vehicle communications.

9.2. SCADA Systems Software

The typical components of a SCADA system, with emphasis on the SCADA software are indicated in the Figure 9.3.

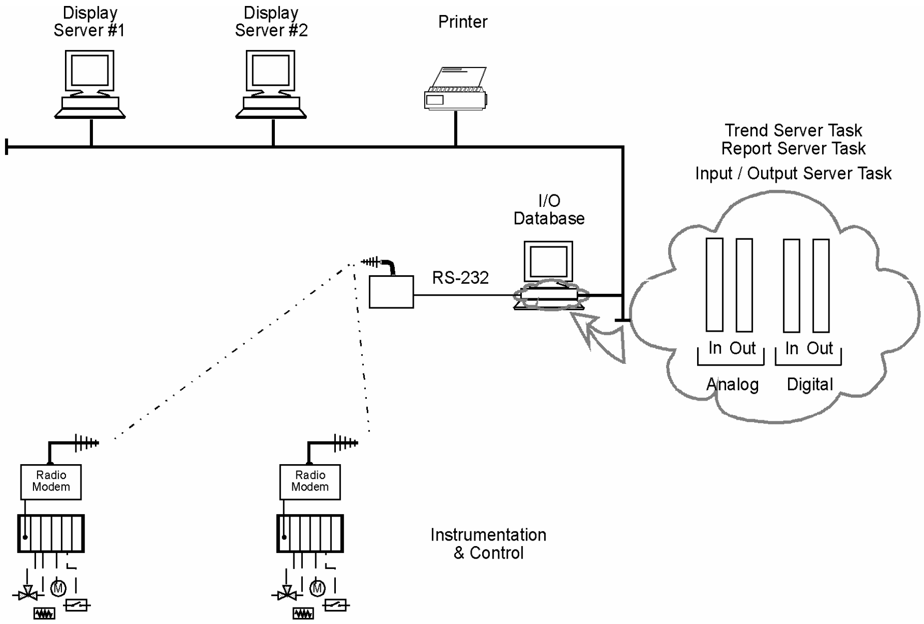


Figure 9.3
Components of a SCADA System

Typical key features expected of the SCADA software are listed below. These features depend on the hardware to be implemented.

9.2.1. SCADA Key Features

User Interface

- Keyboard
- Mouse
- Trackball
- Touch screen

Graphics Displays

- Customer-configurable, object orientated and bit mapped
- Unlimited number of pages
- Resolution: up to 1280 x 1024 with millions of colors

Alarms

- Client server architecture
- Time stamped alarms to 1 millisecond precision (or better)
- Single network Acknowledgment and control of alarms
- Alarms shared to all clients
- Alarms displayed in chronological order
- Dynamic allocation of alarm pages
- User-defined formats and colors

- Up to four adjustable trip points for each analog alarm
- Deviation and rate of change monitoring for analog alarms
- Selective display of alarms by category (256 categories)
- Historical alarm and event logging
- Context-sensitive help
- On-line alarm disable and threshold modification
- Event-triggered alarms
- Alarm-triggered reports
- Operator comments that can be attached to alarms

Trends

- Client server architecture
- True trend printouts (not screen dumps)
- Rubber band trend zooming
- Export data to DBF, CSV files
- X/Y plot capability
- Event based trends
- Pop-up trend display
- Trend gridlines or profiles
- Background trend graphics
- Real-time multi-pen trending
- Short and long term trend display
- Length of data storage and frequency of monitoring that can be specified on a per-point basis
- Archiving of historical trend data
- On-line change of time-base without loss of data
- On-line retrieval of archived historical trend data
- Exact value and time that can be displayed
- Trend data that can be graphically represented in real time

RTU (and PLC) Interface

- All compatible protocols included as standard
- DDE drivers supported
- Interface also possible for RTUs, loop controllers, bar code readers and other equipment
- Driver toolkit available
- Operates on a demand basis instead of the conventional predefined scan method
- Optimization of block data requests to PLCs
- Rationalization of network user data requests
- Maximization of PLC highway bandwidth

Scalability

Additional hardware can be added without replacing or modifying existing equipment. This is limited only by the PLC architecture (typically 300 to 40,000 points)

Access to Data

- Direct, real-time access to data by any network user
- Third-party access to real-time data, e.g. Lotus 123 and EXCEL
- Network DDE
- DDE compatibility: read, write and exec
- DDE to all IO device points
- Clipboard

Database

- ODBC driver support
- Direct SQL commands or high level reporting

Networking

- Supports all NetBIOS compatible networks such as NetWare, LAN Manager, Windows for Workgroups, Windows NT (changed from existing NT)
- Support protocols NetBEUI, IPX/SPX, TCP/IP and more
- Centralized alarm, trend and report processing - data available from anywhere in the network
- Dual networks for full LAN redundancy
- No network configuration required (transparent)
- May be enabled via single check box, no configuration
- LAN licensing based on the number of users logged onto the network, not the number of nodes on the network
- No file server required
- Multi-user system, full communication between operators
- RAS and WAN supported with high performance
- PSTN dial up support

Fault Tolerance and Redundancy

- Dual networks for full LAN redundancy
- Redundancy that can be applied to specific hardware
- Supports primary and secondary equipment configurations
- Intelligent redundancy allows secondary equipment to contribute to processing load
- Automatic changeover and recovery
- Redundant writes to PLCs with no configuration
- Mirrored disk I/O devices
- Mirrored alarm servers
- Mirrored trend servers
- File server redundancy
- No configuration required, may be enabled via single check box, no configuration

Client/Server Distributed Processing

- Open architecture design
- Real-time multitasking
- Client/server fully supported with no user configuration
- Distributed project updates (changes reflected across network)

- Concurrent support of multiple display nodes
- Access any tag from any node
- Access any data (trend, alarm, report) from any node

9.2.2. The SCADA Software Package

Whilst performance and efficiency of the SCADA package with the current plant is important, the package should be easily upgradeable to handle future requirement. The system must be easily modifiable to allow for the requirements changing and expanding as the task grows - in other words the system must use a scaleable architecture.

There have been two main approaches to follow in designing the SCADA system:

- Centralized, where a single computer or mainframe performs all plant monitoring and all plant data is stored on one database which resides on this computer.
- Distributed, where the SCADA system is shared across several small computers (usually PCs).

An effective solution is to examine the type of data required for each task and then to structure the system appropriately. A client server approach also makes for a more effective system.

There are typically five tasks in any SCADA system. Each of these tasks performs its own separate processing.

- Input/Output Task. This program is the interface between the control and monitoring system and the plant floor.
- Alarm Task. This manages all alarms by detecting digital alarm points and comparing the values of analog alarm points to alarm thresholds.
- Trends Task. The trends task collects data to be monitored over time.
- Reports Task. Reports are produced from plant data. These reports can be periodic, event triggered or activated by the operator.
- Display Task. This manages all data to be monitored by the operator and all control actions requested by the operator.

9.2.3. System Response Times

These should be carefully specified for the following events. Typical speeds which are considered acceptable are:

- Display of analogue or digital value (acquired from RTU) on the Master Station Operator Display (1 to 2 seconds maximum)
- Control request from operator to RTU (1 second critical; 3 seconds non- critical)
- Acknowledge of alarm on operator screen (1 second)
- Display of entire new display on operator screen (1 second)
- Retrieval of historical trend and display on operator screen (2 seconds)
- Sequence of events logging (at RTU) of critical events (1 millisecond)

It is important that the response is consistent over all activities of the SCADA system.

9.2.4. Specialized SCADA Protocols

A Protocol controls the message format common to all devices on a network. Common protocols used in radio communications and telemetry systems include the HDLC, MPT1317 and Modbus protocols. The CSMA/CD protocol format is also used.

9.2.4.1. Introduction to Protocols

The transmission of information (both directions) between the master station and RTUs using time division multiplexing techniques requires the use of serial digital messages. These messages must be efficient, secure, flexible, and easily implemented in hardware and software. Efficiency is defined as:

Information Bits Transmitted ÷ Total Bits Transmitted

Security is the ability to detect errors in the original information transmitted, caused by noise on the communication channel. Flexibility allows different amounts and types of information to be transmitted upon command by the master station. Implementation in hardware and software requires the minimum in complicated logic, memory storage, and speed of operation.

All messages are divided into three basic parts as follows:

- Message Establishment; which provides the signals to synchronize the receiver and transmitter.
- Information; which provides the data in a coded form to allow the receiver to decode the information and properly utilize it.
- Message Termination; which provides the message security checks and a means of denoting the end of the message. Message security checks consist of logical operations on the data which result in a predefined number of check bits transmitted with the message. At the receiver the same operations are performed on the data and compared with the received check bits. If they are identical, the message is accepted; otherwise, a retransmission of the original message is requested.

A typical example of commonly used asynchronous message format is shown in Figure 9.4.

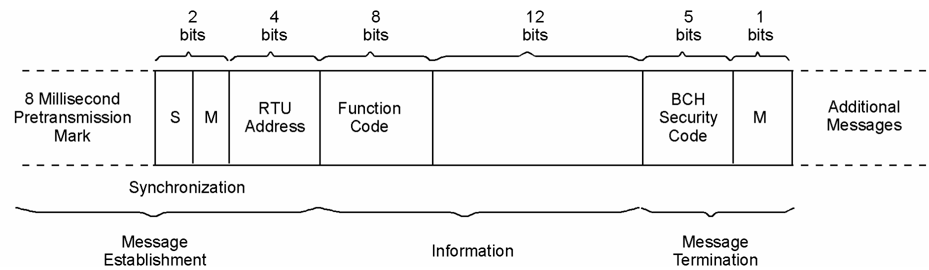


Figure 9.4
Typical Asynchronous Message Format

9.2.4.2. Information Transfer

Master to Remote Data Transfer: Information transmitted from master to remote is for the purpose of device control, set point control, or batch data transfer. Due to the possible severe consequences of operating the wrong device or receiving a bad control message, additional security is required for control. This is provided in the form of a sequence of messages, commonly called a select-before-operate sequence, as shown in Figure 9.5.

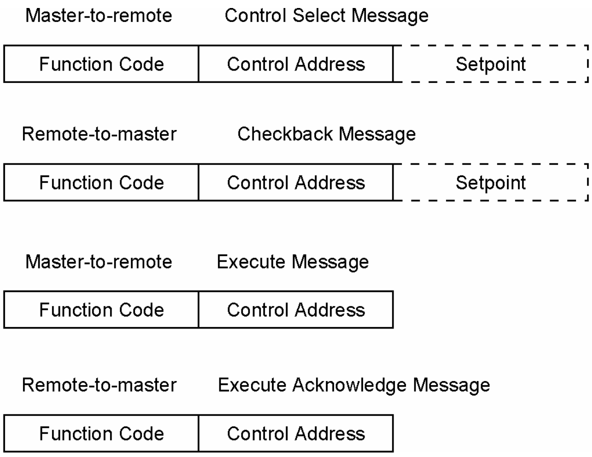


Figure 9.5
Sequence of Messages for Control

The following explanatory notes apply to Figure 9.5:

- Message establishment and message termination fields are not shown
- Function code specifies the operation to be performed by the RTU.
- Control address specifies the device or set point to be controlled
- Set point provides the value to be accepted by the RTU
- A remote to master checkback message is derived from the RTU point selection hardware in order to verify that the RTU has acted correctly in interpreting the control selection.

Remote to Master Data Transfer: All remote to master data transfer is accomplished with one basic message sequence by using variations in the field definitions to accommodate different types of data. The basic sequence is shown in Figure 9. 6.

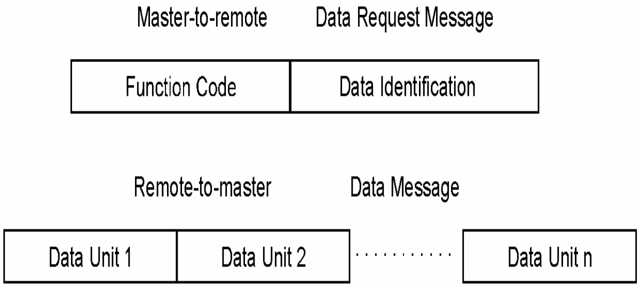


Figure 9. 6
Sequence of Messages for Data Acquisition

The following explanatory notes apply to Figure 9. 6:

- Message establishment and message termination fields are not shown.
- Function code specifies the type of data to be transferred by the RTU.
- Data identification identifies the amount and type of data requested by the master station.

9.2.4.3. High Level Data Link Control (HDLC) Protocol

HDLC has been defined by the International Standards Organization for use on both multipoint and point-to-point links. HDLC is a bit based protocol. The two most common modes of operation of HDLC are:

Unbalanced Normal Response Mode (NRM): This is used with only one primary (or master) station initiating all transactions.

Asynchronous Balanced Mode (ABM): In this mode each node has equal status and can act as either a secondary or primary node.

The standard format is indicated in Figure 9.7 below. .

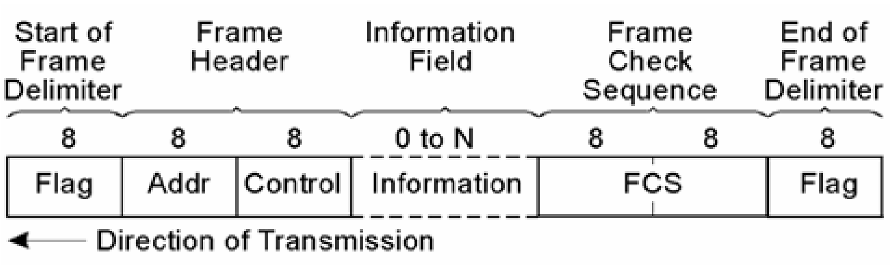


Figure 9.7
HDLC Frame Format

Contents of Frame: The contents of the frame are briefly as follows:

The flag character is a byte with the form 01111110. In order to ensure that the receiver always knows that the character it receives is a unique flag character (rather than merely some other character in the sequence); a procedure called zero insertion is followed. This requires the transmitter to insert a '0' after a sequence of five 1's in the text (i.e. non flag characters).

The Frame Check Sequence (FCS) uses the CRC-CCITT methodology except that 16 ones are added to the tail of the message before the CRC calculation proceeds and the remainder is inverted.

The address field can contain one of three types for the request or response message to or from the secondary node:

- Standard secondary address
- Group addresses for groups of nodes on the network
- Broadcast addresses for all nodes on the network (here the address contains all 1s)

Where there are a large number of secondarys on the network, the address field can be extended beyond 8 bits by encoding the least significant bit as a 1. This then indicates that there is another byte to follow in the address field.

The control field is indicated in Figure 9.7. Note that the send and receive sequence numbers are important to detect and correct errors in the messages. The P/F bit is the poll/final bit and when set indicates to the receiver that it must respond or acknowledge this frame (again with the P/F bit set to 1).

Protocol Operation: A typical sequence of operations is given below.

- In a multidrop link, a normal response mode frame is sent by the primary node with the P/F bit set to 1 together with the address of the secondary.
- The secondary responds with an unnumbered acknowledgment with the P/F bit set to 1. Alternatively if the receiving node is unable to accept the set up command a disconnected mode frame is returned.
- Data is then transferred with the information frames.
- The primary node then sends an unnumbered frame containing disconnect in the control field.
- The secondary then responds with an unnumbered acknowledgment.
- A similar approach is followed for a point to point link using asynchronous balanced mode except that both nodes can initiate the setting up of the link and the transfer of information frames, and the clearing of the point to point link.
- When the secondary transfers the data, it transmits the data as a sequence of information frames with the F bit set to 1 in the final frame of the sequence.
- In NRM mode if the secondary has no further data to transfer, it responds with a receiver not ready frame with the P/F bit set to 1.

9.2.4.4. The CSMA/CD Protocol Format

The CSMA/CD protocol is not as comprehensive as HDLC and is concerned with the method used to get data on and off the physical medium. HDLC and CSMA/CD can be incorporated together for a more complete protocol.

The format of a CSMA/CD frame which is transmitted is shown in Table 9. 1 The MAC frame consists of seven bytes of preamble, one byte of the start frame Delimiter and a data frame. The data frame consists of a 48 bit source and destination address, 16 bits of length or type fields, data and a 32 bit CRC field.

The minimum and maximum sizes of the data frames are 64 bytes and 1518 bytes respectively.

Table 9. 1

Format of a Typical CSMA/CD Frame

Preamble	**SFD**	**Destination Address**	**Source Address**	**Length Indicator**	**Data**	**Frame Check Sequence**
7 Bytes	1 Byte	2 or 6 Bytes	2 or 6 Bytes	2 Bytes		4 Bytes

The format of the frame can be briefly described as follows (with reference to each of the fields): The following sequence is followed for the transmission and reception of a frame.

9.2.5. Distributed Network Protocol

The Distributed Network Protocol is a data acquisition protocol used mostly in the electrical and utility industries. It is designed as an open, interoperable and simple protocol specifically for SCADA controls systems. It uses the master/slave polling method to send and receive information, but also employs sub-masters within the same system. The physical layer is generally designed around RS232 (V.24), but it also supports other physical standards such as RS422, RS 485 and even Fiber Optic.

The DNP is well developed as a device protocol within a complete SCADA system. It is designed as a data acquisition protocol with smart devices in mind. These devices can be coupled as a multi-drop fieldbus system. The fieldbus DNP devices are integrated into a software package to become a SCADA system. DNP does not specify a single physical layer for the Serial bus (multi-mode) topology. Devices can be connected by 422 (four wire), 485 (two wire), modem (Bell 202) or with fiber optic cable. The application program can integrate DNP with other protocols if the SCADA software permits. Using tunneling or encapsulation the DNP could be connected to an Intranet or the Internet.

9.2.6. New Technologies in SCADA Systems

A few of the new developments that are occurring in SCADA technology will be briefly listed below. The rapid advances in communications technology are an important driving force in the new SCADA system.

- Rapid Improvement in LAN Technology for Master Stations
- Man Machine Interface
- Remote Terminal Units
- Communications

9.3. Distributed control system (DCS)

SCADA technology has existed since the early sixties and there are now two other competing approaches possible - Distributed control system (DCS) and Programmable logic controller (PLC).

In a DCS, the data acquisition and control functions are performed by a number of distributed microprocessor-based units, situated near to the devices being controlled or, the instrument from which data is being gathered. DCS systems have evolved into providing very sophisticated analogue (e.g. loop) control capability. A closely integrated set of operator interfaces (or man machine interfaces) is provided to allow for easy system configurations and operator control. The data highway is normally capable of high speeds - typically 1 Mbps up to 10 Mbps (see Figure 9.8).

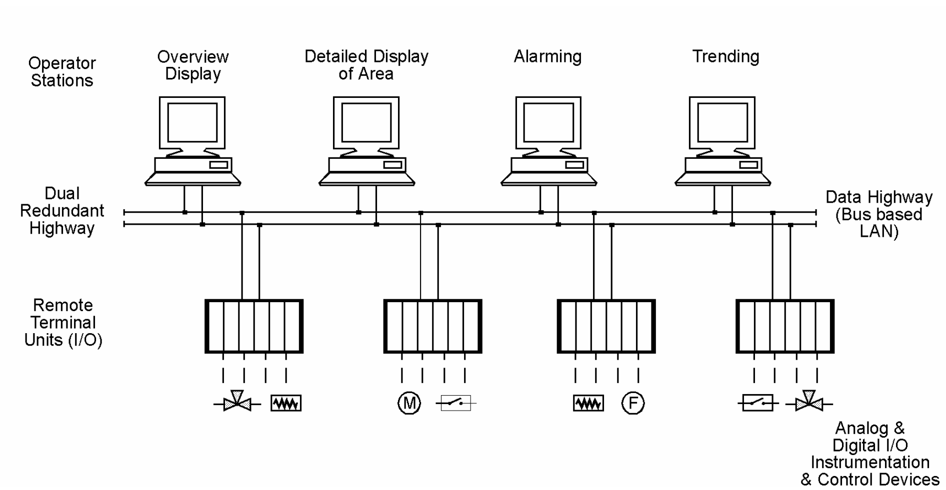


Figure 9.8
Distributed control system (DCS)

9.3.1. DCS versus SCADA terminology

The goals of a DCS (distributed control system) and SCADA (supervisory control and data acquisition system) can be quite different.

A DCS is a process-oriented system and it treats the control of the process, (the chemical plant, refinery or whatever) as its main task, and it presents data to operators as part of its job. On the other hand, a SCADA system is data gathering oriented; and the control center and operators are its focus. Interestingly enough, the remote equipment is merely there to collect the data - though it may also do some very complex process control.

A DCS operator station is intimately connected with its input/output signals (I/O) through local wiring, communication buses (e.g. Field Bus, networks) etc. When the DCS operator wants to see information he/she usually makes a request directly to the field I/O and gets a response. Field events can directly interrupt the system and advise the operator.

A SCADA system must continue to operate when field communications have failed. The 'quality' of data shown to the operator is an important facet of SCADA system operation. SCADA systems often provide special 'event' processing mechanisms to handle conditions that occur between data acquisition periods.

There are many other differences, but they tend to involve a lot of detail. The underlying points are:

- A SCADA system needs to transfer secure data and control signals over a potentially slow, unreliable communications medium, and needs to maintain a database of 'last known good values' for prompt operator display. It frequently needs to do event processing and data quality validation. Redundancy is usually handled in a distributed manner.
- A DCS is always connected to its data source, so it does not need to maintain a database of 'current values'. Redundancy is usually

handled by parallel equipment, not by diffusion of information around a distributed database.

9.3.2. DCS Controller

A DCS controller is a high-performance device capable of handling hundreds of discrete or regulatory control loops per second. The control performance capability of a DCS controller varies from manufacturer to manufacturer. But a user can customize his control configuration to meet the application requirements.

For configuring a DCS controller, it is important to understand the types of controller slots and control functions and algorithms the DCS controller offers.

Control modes: A basic DCS controller has thefollowing operating modes:

- Manual mode
- Automatic mode.
- Cascade mode.
- Backup cascade mode.

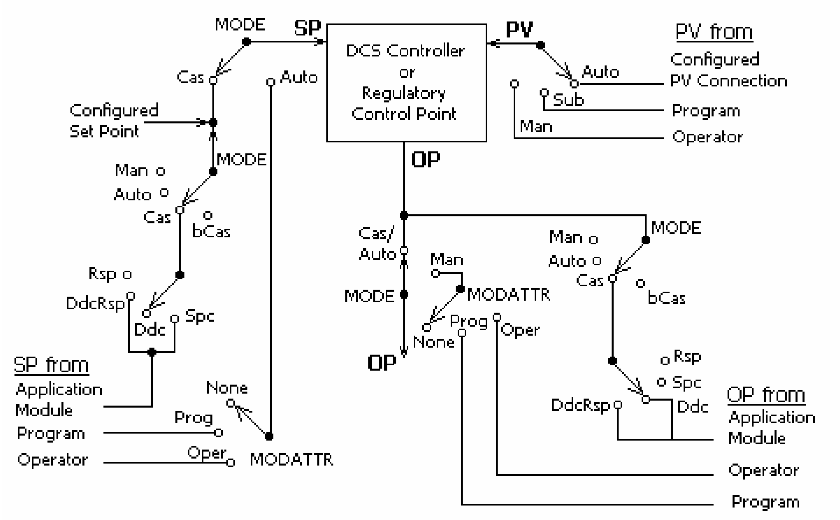


Figure 9.9
Typical mode structure of a basic DCS controller

9.3.3. Control functions

The DCS controller provides a variety of control tools that can be customized to address a wide range of process automation needs.

Functions from I/O scanning through regulatory and logic control to more advanced control strategies can be easily implemented through the DCS advanced controller. Control strategies include; a sophisticated regulatory control package, fully integrated interlock logic functions, and an advanced high level, process engineer-oriented control programming language.

Conceptually, a DCS controller can be thought of as partitioned into ‘slots’ of various types. These slots provide an allocated resource of processing power and money that can be user-configured, including the assignment of a tag name.

A tagged slot is referred to as a ‘data point’ or ‘point’ in some DCS systems. Predefined groups, detail displays as well as custom graphics support this data point structure.
Following are some of the different types of data points that can be configured into a DCS controller slot:

- Regulatory PV
- Regulatory control
- Digital composite
- Logic
- Device control
- Array
- Flag
- Numeric
- Timer
- String, etc.

9.3.4. Control algorithms

For different DCS system controllers, different control algorithms are available. Some common control algorithms follow.

- Proportional, integral, derivative (PID): The PID algorithm operates as a 3-mode (proportional, integral and derivative) controller. One can choose from one of the two forms of this algorithm; the interactive (real) form, and the non-interactive (ideal) form.
- PID control algorithm equations for the interactive form
- PID control algorithm equations for the non-interactive form
- PID with feed-forward (PIDFF) control algorithm
- PID with external reset-feedback (PIDERFB)
- Position proportional control algorithm

9.4. Introduction to the PLC

“PLC” means “Programmable Logic Controller”. The word “Programmable” differentiates it from the conventional hard-wired relay logic. It can be easily programmed or changed as per the application’s requirement. The PLC also surpassed the hazard of changing the wiring.

The PLC as a unit consists of a processor to execute the control action on the field data provided by input and output modules.

In a programming device, the PLC control logic is first developed and then transferred to the PLC.

9.4.1. What can a PLC do?

- It can perform relay-switching tasks.
- It can conduct counting, calculation and comparison of analog process values.
- It offers flexibility to modify the control logic, whenever required, in the shortest time.

- It responds to the changes in process parameters within fractions of seconds.
- It improves the overall control system reliability.
- It is cost effective for controlling complex systems.
- It trouble-shoots more simply and more quickly
- It can be worked with the help of the HMI (Human-Machine Interface) computer

9.4.2. Basic block diagram of the PLC

Figure 9.10 shows the basic block diagram of a common PLC system.

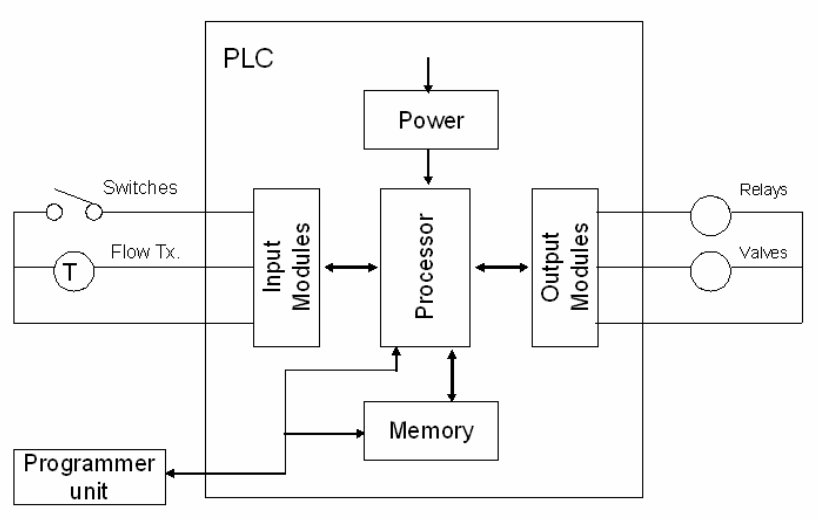


Figure 9.10
Block diagram of a PLC

As shown in the above figure, the heart of the "PLC" in the center, i.e., the Processor or CPU (Central Processing Unit).

- The CPU regulates the PLC program, data storage, and data exchange with I//O modules.
- Input and output modules are the media for data exchange between field devices and CPU. It tells CPU the exact status of field devices and also acts as a tool to control them.
- A programming device is a computer loaded with programming software, which allows a user to create, transfer and make changes in the PLC software.
- Memory provides the storage media for the PLC program as well as for different data.

9.4.3. Size of the PLC system

PLCs are classified on the basis of their size:

- A small system is one with less than 500 analog and digital I/Os.
- A medium system has I/Os ranging from 500 to 5,000.
- A system with over 5,000 I/Os are considered large.

9.4.4. Components of the PLC system

CPU or processor: The main processor (Central Processing Unit or CPU) is a microprocessor-based system that executes the control program after reading the status of field inputs and then sends commands to field outputs.

I/O section: I/O modules act as “Real Data Interface” between field and PLC CPU. The PLC knows the real status of field devices, and controls the field devices by means of the relevant I/O cards.

Programming device: A CPU card can be connected with a programming device through a communication link via a programming port on the CPU.

Operating station: An operating station is commonly used to provide an "Operating Window" to the process. It is usually a separate device (generally a PC), loaded with HMI (Human Machine Software).

9.5. Considerations and benefits of SCADA system

Typical considerations when putting a SCADA system together are:

Overall control requirements

- Sequence logic
- Analog loop control
- Ratio and number of analog to digital points
- Speed of control and data acquisition

Master/Operator control stations

- Type of displays required
- Historical archiving requirements

System consideration

- Reliability/availability
- Speed of communications/update time/system scan rates
- System redundancy
- Expansion capability
- Application software and modeling

Obviously a SCADA system's initial cost has to be justified. A few typical reasons for implementing a SCADA system are:

- Improved operation of the plant or process resulting in savings due to optimization of the system
- Increased productivity of the personnel
- Improved safety of the system due to better information and improved control
- Protection of the plant equipment
- Safeguarding the environment from a failure of the system
- Improved energy savings due to optimization of the plant
- Improved and quicker receipt of data so that clients can be invoiced more quickly and accurately

- Government regulations for safety and metering of gas (for royalties and tax etc).

9.6. An alarm system

In industrial plants and installations, control systems are used to monitor and control processes. Control Systems, whether a conventional Control Desk or a Computer/PLCs System with SCADA or a Distributed Control System (DCS), provides a man-machine-interface to monitor and control the plant equipment and processes.

Alarm Systems are an integral part of man-machine interface. An alarm system consists of both hardware and software including; field signal sensors, transmitters, alarm generators & handlers, alarm processors, alarm displays, annunciator window panels, alarm recorders and printers. Alarm systems indicate the abnormal conditions and problems of the plant and equipment to the operators, enabling them to take corrective action and bring the plant/equipment back to normal conditions. Alarm systems give signals to the operators in the form of audible sound, visual indications in different colors and/or continuous blinking, text messages, etc.

An alarm system brings the following to the notice of the operator:

- problems that need operator attention
- process changes that require corrective action
- unsafe operating conditions before Emergency Shut-down of the plant
- hazardous conditions
- deviations from desired/normal conditions

9.6.1. Functions of the plant or process operator

An alarm system helps/assists the operators in monitoring and controlling the plant, equipment and processes within safe and normal operating conditions. In order to design a suitable alarm system, it is important to understand the functions of the operator who monitors and controls the equipment and processes in the plant.

Generally, the functions of a plant operator are inclusive of the following activities but are not limited to:

- safe and normal operation of plant/equipment
- production at optimum levels
- identification of abnormal, hazardous and unsafe plant/equipment conditions and taking corrective action
- fault identification and communication of faults to maintenance

The above mentioned function and task priorities of a plant operator change with the changing conditions of the plant. For instance:

- during start-up
- when the plant is being stabilized
- when the plant is running under normal conditions
- when the plant is running in abnormal conditions

- when the plant is in emergency shut-down
- when the plant is in planned shut-down,
- when the plant, or sub-section of plant, is in manual mode of operation
- during automatic mode of operation

9.6.2. Functions of an alarm system

The main function of an alarm system is to direct the attention of an operator towards the plant abnormal conditions that need timely assessment and/or timely corrective action(s). An Alarm system alerts, informs and guides an operator regarding an abnormal situation and helps him to take timely corrective action to bring back the plant to normal conditions.

When an abnormal condition arises, the alarm system gives an alarm in the form of an audible warning, flashing or blinking alarm indication and an alarm message. The Alarm gives information about the problem or abnormal condition and its details. In a good alarm system, guidance or help messages on how to respond and take corrections are also provided. An ideal Alarm system also provides feedback on the corrective actions taken by the operator in response to the alarm. Such feedback is generally provided on supplementary display screens that can be accessed by selecting an alarm in the Alarm list.

9.6.3. An effective alarm system

For designing an effective Alarm system, it is important to consider the following key points:

- Present only relevant and useful alarms to the operator
- Each alarm should have a defined response from the operator
- Configure and present only a good alarm
- Allow adequate time for an operator to respond to an alarm

9.6.4. Alarm system design

Designing an alarm system is a process. While designing each alarm it is important to consider how important the alarm is and what its reliability should be. To determine the importance and reliability of an alarm, it is necessary to carry out a qualitative and quantitative risk assessment to consider whether the alarm is safety related and whether it is to be implemented on an independent stand-alone system as opposed to the process control system. Safety related alarms should be given special considerations while designing the man-machine interface.

9.6.5. Assessment of risk

Risk is a measure of the probable rate of occurrence of a hazard and its severity. Risk can be applied to safety hazards, environmental hazards and economic losses.

Alarms are configured and presented to the operator to take corrective action(s) and minimize the sub-optimal operations of the plant/equipment or to protect plant/equipment from damages that can lead to injury to people, damage to environment and/or economic losses. So the design of an alarm system should consider these risks and it must be clearly identified which risk is intended to be reduced by the alarm.

9.6.6. Protection provided by the alarm system

Protection provided by an alarm system can take place in two ways. The operator is warned by the alarm and he/she takes corrective action before the protection operates, or the operator is warned that the protection has failed to operate and he/she takes corrective action.

9.6.7. Safety related alarms

As per the international standard IEC 61508, an alarm system, whether electrical or electronic or programmable, should be considered as safety related only if:

- It is a claimed part of the facilities for reducing the risk(s) from hazards to people to an acceptable or tolerable level, and
- The claimed reduction by the alarm system in the risk(s) is significant. Here the significant reduction means a claimed Average Probability of Failure on Demand (PFDAvg) < 0.1,
- It is designed, operated and maintained as per the requirements defined in the standard,
- It is independent and separate from the process control system, unless the process control system itself has been identified as a safety related system and implemented accordingly.

9.6.8. What is the purpose of an alarm?

(i) It is important to know what the purpose is of the proposed alarm and for what hazards or risks it will provide a warning or an alert to the operator. The consequences of alarm failure or the alarm being missed need to be identified. If the proposed alarm provides only information of an event/incident, then it should not be configured as an alarm.

(ii) Assessment of the severity of the risk in terms of potential loss of life or an injury, economic losses, environmental impact and plant damages must be done. Any hazard to people should be in the form of formal risk assessment for the plant. Economic risks, potential plant damages or losses should be expressed in terms of financial losses.

(iii) Expected frequency of the risk occurrence should be estimated. Though it is difficult to know the accurate chances/frequency of occurrence, it may be appropriate to have some approximate estimate that is more realistic. Appropriate frequency of occurrence may be specified as once a week or once in month, etc.

(iv) Are there any other protection systems in the plant to provide protection against the risk? If not, then it needs to be decided whether or not an automatic protective system can be used with or without configuring the alarm.

(v) Are any reliability claims made in the plant, in terms of safety and protection, provided by the alarm? Do these reliability claims require the alarm to be classified as a safety related alarm? If an alarm is not safety related, then what are the economic and/or environmental risks involved in implementing the alarm within the process control system?

(vi) It is important to know the implications of alarm failure due to alarm sensor/instrument failure. How then can these failures be detected and can the alarm signal be validated. Should the alarm sensor/instrument be made redundant?

(vii) How effective will the operator response to the alarm be? If the operator cannot take any corrective or preventive action to prevent the risk, then the alarm hardly provides any benefit and should not be configured as an alarm.

9.6.9. Operator response

(i) What should the response of the operator be to the alarm? The response may be an action, a conditional action or only a cognitive switch. The response must be clearly defined for each alarm.

(ii) The alarm message should be easy to read and understand.

(iii) If required, additional displays should be developed that provide the operator with information to help him decide how to respond in different conditions of the plant.

(iv) How long will the operator take to respond to the alarm and how long will the plant take to respond to the corrective action(s) taken by the operator?

(v) Procedure for Alarm response should be prepared and provided to the operators.

9.6.10. Alarm prioritization

(i) The likely safety, economic and environmental consequences of the operator not responding to the alarm should be assessed.

(ii) Identify whether the alarm is time critical and what time is available for the operator to respond to the alarm.

(iii) Depending on the severity of safety, economic and environmental consequences of missing the alarm and/or not responding to a time critical alarm and how fast the operator is required to respond to the alarm, priority should be allocated to the alarm.

(iv) It should be determined whether it will be required to change the priority of the alarm depending on changes in the operating conditions.

9.6.11. Alarm settings

(i) What is the normal range for the alarmed process variable? What should be the alarm settings for alarming the safety hazard, and/or economic losses and/or environmental damages?

(ii) How many alarms should be set for the process variable – Low, Low-Low, High, High-High and what should the settings be for these alarms?

(iii) Is there a need for changing the alarm setting depending on plant operating conditions?

(iv) During normal plant operation, what are the fluctuations in the process variable to be alarmed?

9.6.12. Alarm suppression

(i) If the alarm is likely to be generated during large disturbance/upset in the plant or during plant trips, then the alarm should be suppressed using a suitable logic.

(ii) If other alarms, having more significance, occur, the alarm should be suppressed?

(iii) Are there any conditions during which the process variable will cross the alarm setting but where no risk is involved?

(iv) What signals should be used for logical suppression of the alarm?

(v) What are the chances of the process variable going out of range or being invalid? How will it affect the alarm? Will it create a nuisance (repeating) alarm? How should such situations be made known to the operator, or how can plant/equipment trip be avoided if the process variable alarm limit/Flag is used for plant/equipment interlock?

Chapter 10. Project Management of I&C Projects

10.1. Fundamentals of project management

Operations and projects share a number of characteristics in that they are:

- Planned, executed, and controlled
- Constrained by resource limitations
- Performed by people

Projects are, however, different from operations (such as maintenance or repair work) in that they are temporary endeavors undertaken to create a unique product or service.

The primary objectives of a project are commonly defined by reference to function, time, and cost.

10.1.1. Project management

Project management is the application of specific knowledge, skills, tools, and techniques to plan, organise, initiate, and control the implementation of the project, in order to achieve the desired outcome(s) safely.

Note that 'Project Management' is also used as a term to describe an organisational approach known as 'Management by Projects', in which elements of ongoing operations are treated as projects, and project management techniques are applied to these elements.

10.1.2. Project life cycle

Projects proceed through a sequence of phases from concept to completion. Collectively, the separate phases comprise the project 'life cycle'.

There are only a limited number of generic lifecycles. The generic types are usually considered to include capital works, pharmaceutical, petrochemical, defence procurement, research and development. Consequently the initial starting point for managing the project is to define the type, and select an appropriate life cycle model as the planning framework.

10.1.3. Project organization

Where projects are set up within existing organizations, the structure and culture of the parent organization has great influence on the project, and will be a deciding factor in whether or not there is a successful outcome. The organization of the project team directly influences the probability of achieving a successful outcome.

The influence of the organization structure on various project parameters is illustrated in Figure 10.1

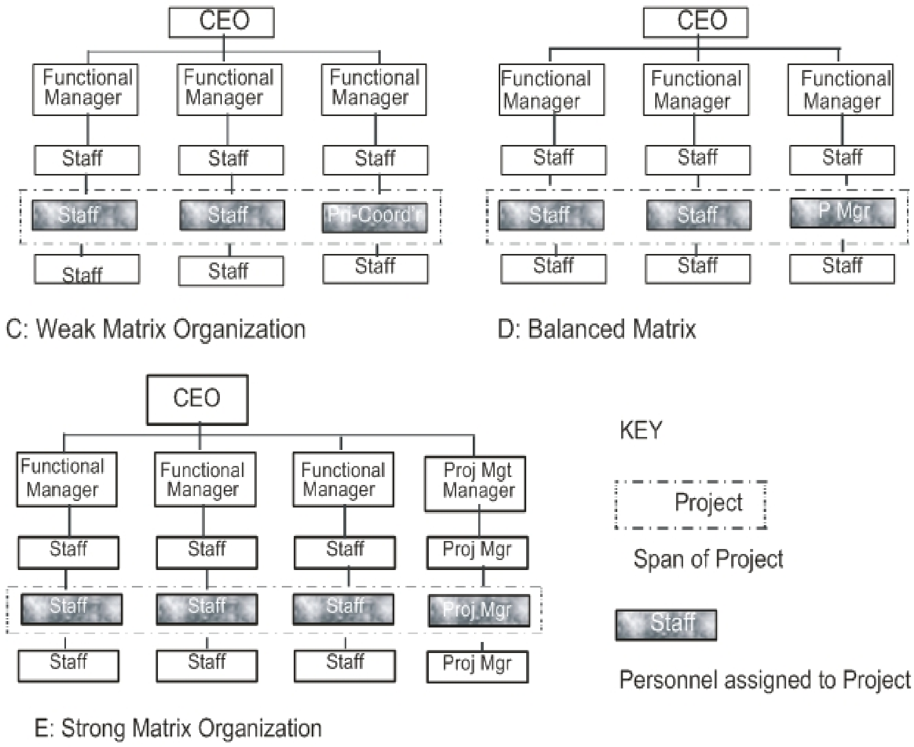


Figure 10.1
Project structures within organizations

The organization of the project team is characterized by:

- The principal or project sponsor.
- The Project Control Group (PCG).
- The project manager. In a 'perfect world' the responsibilities, roles and authority of this person would be defined and documented.
- A project control officer or group, if this function is not undertaken by the project manager.
- The rest of the project team, which will vary in composition according to the project type, as well as specific project variables.

10.1.4. Project planning

The project planning phase is critical to the effective implementation and control of the project and the basis for project success is established during this phase. The primary output from this phase is the Project Quality Plan (PQP). The basic element required to properly define the PQP is the Work Breakdown Structure or WBS.

The PQP comprises the following:

- The PQP sign-off
- The statement of project objectives
- The project charter
- The project plan
- Project control procedures

10.2. Time management

Time management of a project consists of:

- Planning the project activities to a time scale (i.e. the project schedule)
- Monitoring performance of the implementation phase
- Comparing achieved performance with the project schedule
- Taking corrective action to ensure planned objectives are most likely to be met.

10.2.1. Project planning

The principal aim of project management is to effectively utilize the available resources in order to achieve the planned objective(s).

The most commonly used methods in the field of project management are known collectively as Project Network Techniques. These comprise:

- The 'Critical Path' method (also known as the Activity on Arrow or AoA method)
- The 'Precedence' method (also known as the Activity on Node or AoN method)

Precedence network analyses are normally presented graphically, either as the network diagram itself, or as a time-scaled bar chart known as a Gantt chart. Critical path networks can be presented as time-scaled arrow diagrams, or as Gantt charts.

Project analysis by either method involves the same four steps:

- Defining the activities, for the initial project plan this may involve the breakdown of work packages used as the basic elements for the other components of the PQP.
- Preparation of the logic sequence to determine the relationships between the activities.
- Applying activity (time and resource) data for each activity.
- Analysis of the network.

10.3. Cost Management

Effective cost management is a key element of successful project management. The critical aspects of the cost management process have to be reviewed, that must be properly addressed to ensure effective financial control.

Cost management includes the processes required to ensure that the project/contract is completed within the approved budget. These processes comprise:

- Cost estimating
- Budgeting
- Financial control
- Change control
- Cost monitoring
- Value management

Cost estimating: Estimating costs is of fundamental importance to every project. Estimates are prepared to meet two different objectives:

- As the basis for determining the economic feasibility of a project
- As the basis for cost management of the project.

Budgeting: Budgeting involves allocating the cost estimates against the project/contract schedule. The budgeting process should establish a cost baseline that provides:

- The basis against which project performance may be measured
- A forecast payment schedule to allow for funds management by the principal

Financial control: The definition of financial controls should extend to:

- Controls over the commitment of funds, i.e. financial authority
- Controls over the approval of expenditure, i.e. authorization of payments

Change control: Effective change control is a vital element of project cost control. This process is often referred to as scope control.

- Following approval of the project budget, there will be unavoidable changes to the project arising from discretionary and non-discretionary sources.

Cost reporting: The basic objective of the financial report must be to provide an accurate status report of forecast financial cost versus approved budget. To meet this objective, the financial report needs to include the following information:

- Initial budget
- Approved budget variances to date
- Current budget
- Current forecast final cost

Value management: The objective of the Value Management (VM) process is to achieve the lowest possible cost without prejudicing required functionality or necessary quality; i.e. to improve the value/cost relationship.

10.4. Integrated cost and time management

The objective of integrated cost and time management is to demonstrate the power of applying Earned Value Analysis (EVA) to measuring and predicting project performance and to develop skill in the application of the Performance Measurement System.

Earned value analysis: Earned Value Analysis (EVA) is the analytical part of what is known as Earned Value Management (EVM), so the two names are often used interchangeably.

This provides a management tool that determines the extent of cost variances that may be separately attributed to over/under expenditure and to scheduled deviations.

A significant benefit of the EVM approach is that most of the data can be presented graphically, in one macroscopic view of the project progress.

10.5. Management of project team

Projects exist to create a unique product or service within a limited time frame. Projects are performed by people. If there is more than one person working in a project then it is called a team. A team is a mixture of a wide assortment of personalities, skills, needs, and issues.

10.5.1. Creating a team culture

Cultures within the project team are a reflection of the project manager's preferences and style.

Project managers need to specifically address the following:

- Clarifying roles, responsibilities and levels of authority for all team members
- Setting up effective channels of communication with all team members
- Setting realistic performance expectations for each team member
- Delegating effectively (instead of over-supervising)
- Publicly rewarding good performance
- Acknowledging legitimate concerns and conflicts within the team

At each stage the project manager can communicate values that move the team to the next stage. The stages in team development are referred to as the processes of ‘forming’, ‘storming’, ‘norming’ and ‘performing’.

10.5.2. Team motivation

There are certain issues that spur on individuals within the team to even greater achievements, and other issues that stifle performance.

The major motivators are (in descending order of importance):

- Achievement
- Recognition
- The work itself

- Responsibility
- Advancement
- Personal growth

10.6. Risk Management

The objective of Risk management is to set out a basis for risk management that will provide sufficient understanding of the process for implementing effective risk management for a specific project.

Definition of 'risk': Risk is the exposure to a process or event that prejudices the successful achievement of the project outcome, by adversely impacting on cost, time, or functional objectives.

The elements of risk are:

- The likelihood of the event arising; and
- The consequences if it does arise

10.6.1. Elements of Risk management

The main elements of the risk management process are:

- Establishing the context
- Risk identification
- Risk analysis
- Risk assessment
- Risk treatment
- Monitoring and reviewing

Establishing the context: The outputs from this step are:

- Definition of the elements within the project to define a structure for the identification and analysis of risks; and
- Definition of risk assessment criteria directly related to the policies, objectives and interests of stakeholders

This process reviews the strategic, organizational and project contexts.

Risk identification: The purpose of this step is to identify all the risks, including those not under the control of the organization, which may impact on the framework defined above.

A systematic process is essential because a risk not identified during this step is removed from further consideration.

Risk analysis: The objectives of risk analysis are to:

- Assign a level of risk to each identified event
- Provide data to assist the assessment and treatment processes
- Separate minor, acceptable risks from others requiring further consideration

Risk is analyzed by consideration of the likelihood and consequence of events occurring within the context of existing controls i.e. management procedures, technical systems and risk management procedures.

Risk assessment: Risk assessment is the process of comparing the levels of risks determined from the analysis process against the acceptance criteria previously established.

The output from the risk assessment is a prioritized list of risks requiring further action.

Risk treatment: Risk treatment involves identifying the range of options available for treating risks identified as requiring action in the previous stage, evaluating those options in respect of each risk, and developing and implementing risk treatment plans.

Note that some risk response activities may have been undertaken during the qualitative analysis step, if the urgency of developing a response to specific risks warranted it.

Monitoring and review: Monitoring and review of all elements of the risk management programme is essential.

The specific risks themselves, as well as the effectiveness of the control measures, need to be monitored. Few risks remain static and factors impacting on the likelihood or consequences may change.

Changing circumstances may alter earlier priorities, and factors impacting on the cost/benefit of management strategies may vary.

10.7. Contract law

The selected issues of contract law are to be reviewed, where a basic understanding of the applicable law is likely to be of direct assistance to personnel involved in the administration of procurement contracts. The issues addressed include:

- The basis of Commonwealth law
- The essential elements of contracts
- Procurement strategies
- Tendering
- Vitiating factors, i.e. those factors that reduce or remove the legal force of the contract
- Termination of contracts
- Extensions of time
- Remedies available for breach of a contract
- Penalties and bonuses

10.7.1. The Commonwealth legal system

In the Commonwealth legal system the Legislature makes the laws, the Executive administers the laws and the Judiciary decides in the case of disputes or transgressions of the laws.

The only way in which rules can be enacted so as to apply generally is by Act of Parliament. Some of Parliament's legislative functions are delegated to subordinate bodies who, within a limited field, are allowed to enact rules.

10.7.2. Elements of contracts

- The contract may be oral, or written, or implied by the conduct of the parties.
- Any simple contract put into writing with adequate particulars becomes a formal contract. It is then a formal simple contract.
- When made by a company it is generally under seal. The period of limitation is 12 years for breach of contracts made under seal. In respect of building construction contracts, S91 of the Building Act 1991 sets the period of limitation at 10 years.
- A contract made by deed must be under seal. If made by deed, no consideration is required, for example, a will or arbitration agreement.

The essential requirements for a contract to be legally enforceable are:

- The intention to create legal relations
- Agreement (offer and acceptance)
- Consideration
- Definite terms
- Legality
- Capacity of the parties

10.7.3. Procurement strategy issues

Consideration of the appropriate strategy should include the following factors:

- Risk preferences of the Principal
- The requirement to demonstrate a competitive process
- The need and/or ability to properly define scope of work prior to commencement
- The need or ability to exercise control over the operations of the Contractor
- The need to achieve early completion

There are a number of separate aspects to be considered when determining construction procurement strategy. These include:

- Tendering strategies
- Pricing strategies
- Timing strategies
- Contract types
- Delivery strategies

10.7.4. Contract types

Construction contracts may be one of the following three basic types, or a combination thereof. Contracts of each type may or may not be subject to cost escalation: if not, they are described as ‘fixed price’ contracts.

Lump sum: Irrespective of the inputs actually necessary to properly complete the works defined in the contract document, the Contractor is entitled to be paid only the contract sum.

10.7.5. Delivery strategies

For service delivery contracts the principal delivery strategies are:

- Defining service requirements based on inputs to be provided.
- Defining service requirements based on performance and condition criteria for serviced units.

10.7.6. Tendering

The law relating to tendering practice is in a process of change as the result of recent case law developments, primarily in Canada. It may be assumed that in due course these cases will be followed throughout the Commonwealth. Previously it was generally regarded, at least by the practitioners, that no contractual obligations arise prior to acceptance of a tender.

This approach does protect a Contractor from the risk of sub-Contractors revoking their tender.

10.7.7. Vitiating factors

Contracts can be:

- Void: The contract is defective in law, a nullity; e.g. an immoral contract
- Avoidable: The contract is binding, but one party has the right, at his option, to set it aside

The following factors make contracts void or avoidable:

- Mistakes
- Misrepresentation
- Duress and undue influence

10.7.8. Termination of contracts

The contract is discharged when the obligation ceases to be binding on the promisor, who is then under no obligation to perform. This arises from:

- Performance
- Agreement
- Passage of time
- Frustration
- Repudiation
- Determination
- Operation of law

10.7.9. Time for completion and extensions of time

- Time for completion
- Provisions to extend the time for completion
- Determination of extensions of time
- Notification of the time extension
- Acceleration
- Payment arising from an extension of time

10.7.10. Remedies for breach of contract

Damages for breach of contract are governed by two considerations, namely the remoteness and the measure of damages.

The contracting parties may agree as a term of the contract the amount to be paid in the event of a specific breach. This sum is either a liquidated damage or a penalty.

The right to an action for damages may be released by:

- Release under seal
- Release by accord and satisfaction
- Fluxion of time, i.e. the right to claim damages is extinguished by the passage of time.

10.7.11. Penalties and bonuses

A penalty is a sum in the nature of a threat to secure performance. Courts will not enforce penalties.

Building contracts may provide for a bonus to be paid, for example for early completion, or for completion below a defined price.

Chapter 11. Latest Instrumentation and Valve Developments

11.1. Basic Measurement performance terms and specifications

There are a number of criteria that must be satisfied when specifying process measurement equipment.

Accuracy: The accuracy specified by a device is the amount of error that may occur when measurements are taken. It determines how precise or correct the measurements are to the actual value and is used to determine the suitability of the measuring equipment.

Accuracy can be expressed as any of the following:

- Error in units of the measured value
- Percent of span
- Percent of upper range value
- Percent of scale length
- Percent of actual output value

Range of operation: The range of operation defines the high and low operating limits between which the device will operate correctly, and at which the other specifications are guaranteed.

Budget/Cost: Although not so much a specification, the cost of the equipment is certainly a selection consideration. This is generally dictated by the budget allocated for the application. Even if all the other specifications are met, this can prove an inhibiting factor.

11.2. Advanced Measurement Performance terms and Specifications

More critical control applications may be affected by different response characteristics. In these circumstances the following may need to be considered:

11.2.1. Hysteresis

Hysteresis is the difference in the output for given input - when the input is increasing and output for same input when input is decreasing. When input of any instrument is slowly varied from zero to full scale and then back to zero, its output varies as shown in Figure 11.1.

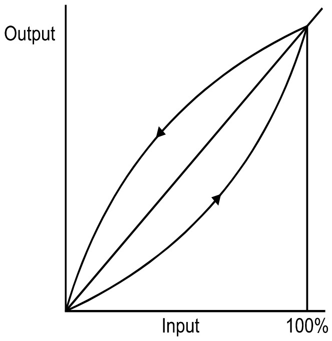


Figure 11.1
Hysteresis.

11.2.2. Linearity

Linearity expresses the deviation of the actual reading from a straight line. If all outputs are in the same proportion to corresponding inputs over a span of values, then input output plot is a straight line else it will be non linear as shown in Figure 11.2. For continuous control applications, the problems arise due to the changes in the rate the output differs from the instrument.

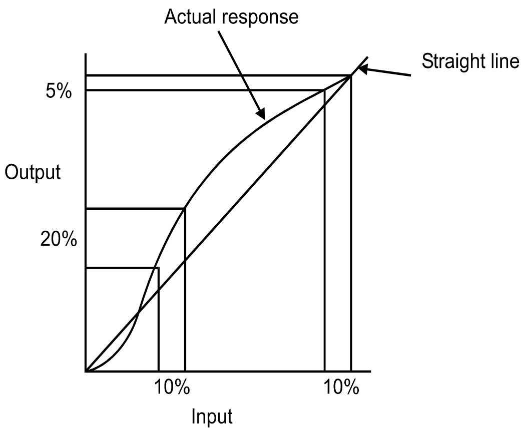


Figure 11.2
Linearity

11.2.3. Repeatability

Repeatability defines how close a second measurement is to the first under the same operating conditions, and for the same input. Repeatability is generally within the accuracy range of a device and is different from hysteresis in that the operating direction and conditions must be the same.

Continuous control applications can be affected by variations due to repeatability. When a control system sees a change in the parameter it is controlling, it will adjust its output accordingly. However if the change is due to the repeatability of the measuring device, then the controller will over-control.

11.2.4. Response

When the output of a device is expressed as a function of time (due to an applied input) the time taken to respond can provide critical information about the suitability of the device. A slow responding device may not be suitable for an application.

11.3. Pressure Measurement

Let us first discuss the principles of pressure measurement

11.3.1. Bar and Pascal

Pressure is defined as a force per unit area, and can be measured in units such as psi (pounds per square inch), inches of water, millimetres of mercury, pascal (Pa, or N/m^2) or bar. Until the introduction of SI units, the 'bar' was quite common.

11.3.2. Absolute, Gauge and Differential pressure

The pascal is a means of measuring a quantity of pressure. When the pressure is measured in reference to an absolute vacuum (no atmospheric conditions), then the result will be in pascal (Absolute). However when the pressure is measured relative to the atmospheric pressure, then the result will be termed pascal (Gauge). If the gauge is used to measure the difference between two pressures, it then becomes Pascal (Differential).

11.3.3. Pressure Sources

Static Pressure: In the atmosphere at any point, static pressure is exerted equally in all directions. Static pressure is the result of the weight of all the air molecules above that point pressing down.

Figure 11.3 shows static pressure acting on hand and static pressure measuring device. Static pressure does not involve the relative movement of the air.

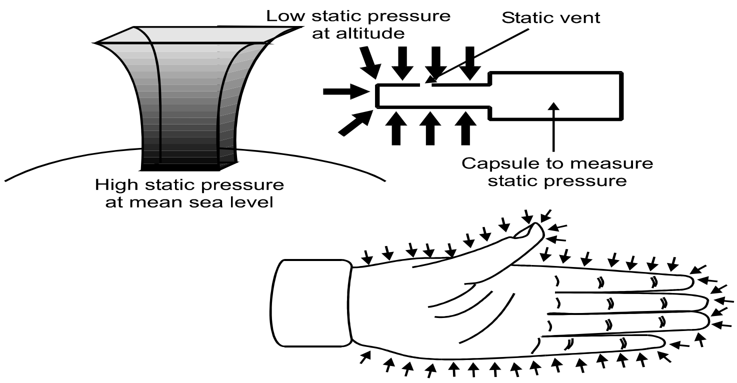


Figure 11.3
Static pressure

Dynamic Pressure: Quite simply, if you hold your hand up in a strong wind or out of the window of a moving car, then the extra wind pressure is felt due to the air impacting your hand.

This extra pressure is over and above the (always-present) static pressure, and is called the dynamic pressure. The dynamic pressure is due to relative movement. Dynamic pressure occurs when a body is moving through the air, or the air is flowing past the body.

Total Pressure: In the atmosphere, some static pressure is always exerted, but for dynamic pressure to be exerted there must be motion of the body relative to the air. Total pressure is the sum of the static pressure and the dynamic pressure.

11.3.4. Pressure Transducers and Elements - Mechanical

- C-Bourdon tube
- Helix and spiral tubes
- Spring and bellows
- Diaphragm
- Manometer
- Single and Double inverted bell

C-Bourdon Tube: The Bourdon tube works on a simple principle that a bent tube will change its shape when exposed to variations of internal and external pressure. As pressure is applied internally, the tube straightens and returns to its original form when the pressure is released.

Helix and Spiral Tubes: Helix and spiral tubes are fabricated from tubing into shapes as per their naming. With one end sealed, the pressure exerted on the tube causes the tube to straighten out. The amount of straightening or uncoiling is determined by the pressure applied.

Spring and Bellows: A bellows is an expandable element and is made up of a series of folds, which allow expansion. One end of the Bellows is fixed and the other moves in response to the applied pressure. A spring is used to oppose the applied force and a linkage connects the end of the bellows to a pointer for indication.

Diaphragm: Many pressure sensors depend on the deflection of a diaphragm for measurement. The diaphragm is a flexible disc, which can be either flat or with concentric corrugations and is made from sheet metal with high tolerance dimensions.

Manometer: The simplest form of a manometer is that of a U-shaped tube filled with liquid. The reference pressure and the pressure to be measured are applied to the open ends of the tube. If there is a difference in pressure, then the heights of the liquid on the two sides of the tube will be different.

Single and Double inverted Bell: The Bell instrument measures the pressure difference in a compartment on each side of a bell-shaped chamber. The bell instrument is used in applications where very low pressures are required to be measured, typically in the order of 0-250 Pa.

11.3.5. Pressure Transducers and Element - Electrical

The typical range of transducers here is:

- Strain gauge
- Vibrating wire
- Piezoelectric
- Capacitance
- Linear Variable Differential Transformer
- Optical
- Strain Gauge

Strain gauge: Strain gauge sensing uses a metal wire or semiconductor chip to measure changes in pressure. A change in pressure causes a change in resistance as the metal is deformed. This deformation is not permanent, as the pressure (applied force) does not exceed the elastic limit of the metal. If the elastic limit is exceeded then permanent deformation will occur.

This is commonly used in a wheatstone bridge arrangement where the change in pressure is detected as a change in the measured voltage. Figure 11.4 shows a simple Wheatstone circuit for strain gauge.

.

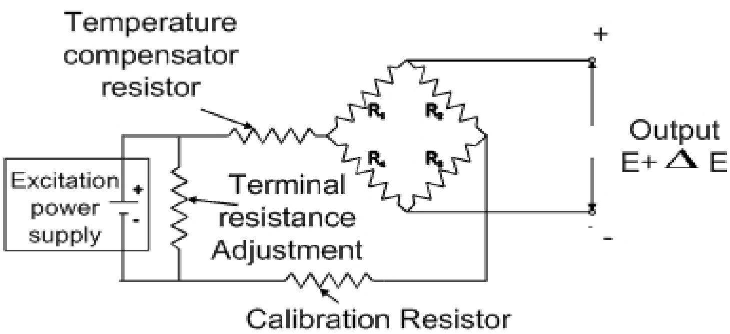


Figure 11.4
Wheatstone circuit for strain gauges

Strain gauges are mainly used due to their small size and fast response to load changes.

Vibrating wire: This type of sensor consists of an electronic oscillator circuit, which causes a wire to vibrate at its natural frequency when under tension. The principle is similar to that of a guitar string. The vibrating wire is located in a diaphragm. As the pressure changes on the diaphragm so does the tension on the wire, which affects the frequency that the wire vibrates or resonates at. These frequency changes are a direct consequence of pressure changes and as such are detected and shown as pressure.

Piezoelectric: When pressure is applied to crystals, they are elastically deformed. Piezoelectric pressure sensing involves the measurement of such deformation. When a crystal is deformed, an electric charge is generated for only a few seconds. The electrical signal is proportional to the applied force.

Because these sensors can only measure for a short period, they are not suitable for static pressure measurement.

Capacitance: Capacitive pressure measurement involves sensing the change in capacitance that results from the movement of a diaphragm. The sensor is energized electrically with a high frequency oscillator.

Linear Variable Differential Transformer: This type of pressure measurement relies on the movement of a high permeability core within transformer coils. The movement is transferred from the process medium to the core by use of a diaphragm, bellows or bourdon tube.

The LVDT operates on the inductance ratio between the coils. Three coils are wound onto the same insulating tube containing the high permeability iron core. The primary coil is located between the two secondary coils and is energized with an alternating current.

Equal voltages are induced in the secondary coils if the core is in the centre. The voltages are induced by the magnetic flux. When the core is moved from the centre position, the result of the voltages in the secondary windings will be different. The secondary coils are usually wired in series. Figure 11.5 shows LVDT for measurement of pressure.

LVDTs are sensitive to vibration and are subject to mechanical wear.

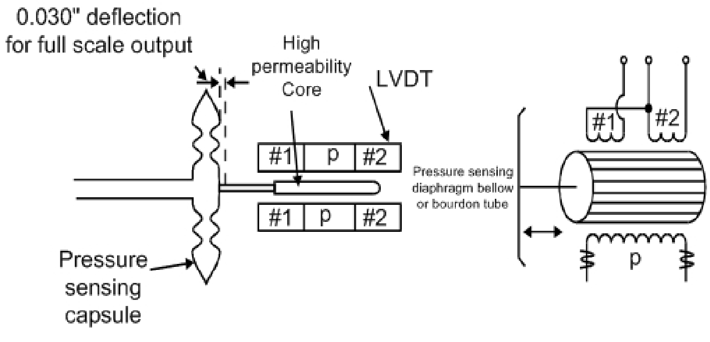


Figure 11.5
Linear variable differential transformer

11.3.6. Installation Consideration

There are a number of points to consider in a pressure measurement application. All require some thought in both the planning and execution.

- Location of Process Connections
- Isolation Valves
- Use of Impulse Tubing
- Test and Drain Valves
- Sensor Construction
- Temperature Effects
- Remote Diaphragm Seals
- Process Flanges
- Additional Hardware

11.4. Level Measurement

11.4.1. Principles of Level Measurement

11.4.1.1. Continuous measurement

The units of level are generally meters (m). However, there are numerous ways to measure level that require different technologies and various units of measurement.

Such means may be:

- Ultrasonic, transit time
- Pulse echo
- Pulse radar
- Pressure, hydrostatic
- Weight, strain gauge
- Conductivity
- Capacitive

For continuous measurement, the level is detected and converted into a signal that is proportional to the level. Microprocessor based devices can indicate level or volume.

11.4.1.2. Point Detection

Point detection can also be provided for all liquids and solids. Some of the more common types are:

- Capacitive
- Microwave
- Radioactive
- Vibration
- Conductive

11.4.2. Simple Sight Glasses and Gauging Rod

11.4.2.1. Simple Sight Glasses

A visual indication of the level can be obtained when part of the vessel is constructed from transparent material or the liquid in a vessel is bypassed through

a transparent tube. The advantage of using stop valves with the use of a bypass pipe is that it facilitates its removal for cleaning.

11.4.2.2. Gauging Rod Method

This requires a little more manual effort than the sight glass, but is another very simple and cheap method of accounting for level. This method can be applied to liquids and bulk materials, and weighted steel tapes can be used in very tall silos.

11.4.3. Buoyancy Tape

There are two main types of buoyancy tape systems available:

Float and Tape Systems: One common form of level measuring system uses a tape or servo motor which is connected to a float. The height can be read as the float moves with liquid level.

Other systems use the float method by sensing the position of the float magnetically or electrically.

- Wire Guided Float Detector:
- For large level measurements (i.e. 20m), wire-guided float detectors can be used.

11.4.4. Hydrostatic Pressure

Some of the different types of level measurement with pressure are:

- Static pressure
- Differential pressure
- Bubble tube method
- Diaphragm Box
- Weighing
- Static Pressure

The basis of hydrostatic pressure measurement for level is such that the measured pressure is proportional to the height of liquid in the tank, irrespective of volume. The pressure is related to the height by the following:

$$P = h.\rho.g$$

where: P = pressure
h = height
ρ = relative density of fluid
g = acceleration due to gravity

For constant density, the only variable that changes is the height. In fact, any instrument that can measure pressure can be calibrated to read height of a given liquid, and can be used to measure liquid level in vessels under atmospheric conditions.

11.5. Temperature Measurement

11.5.1. Principles of Temperature Measurement

Temperature measurement relies on the transfer of heat energy from the process material to the measuring device. The measuring device therefore needs to be temperature dependent.

There are two main industrial types of temperature sensors:

- Contact
- Non contact

Contact: Contact is the more common and widely used form of temperature measurement. The three main types are:

- Thermocouples
- Resistance Temperature Detectors (RTD's)
- Thermistors

These types of temperature devices all vary in electrical resistance for temperature change. The rate and proportion of change is different between the three types, and also different within the type classes.

Non-Contact: Temperature measurement by non-contact means is more specialized and can be performed with the following technologies:

- Infrared
- Acoustic

11.6. Thermocouples

A Thermocouple consists of two wires of dissimilar metals, such as iron and constantan, electrically connected at one end as shown in Figure 11.6. Applying heat to the junction of the two metals produces a voltage between the two wires. This voltage is called an emf (electro-motive force) and is proportional to temperature.

A thermocouple requires a reference junction, placed in series with the sensing junction. As the two junctions are at different temperatures a thermal emf is generated. The reference junction is used to correct the sensing junction measurement.

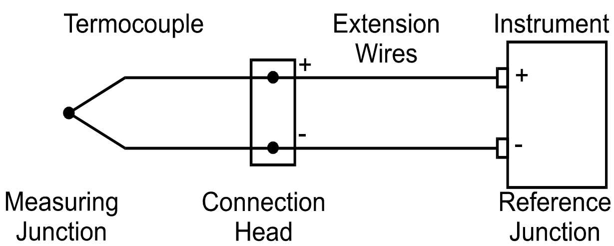


Figure 11.6
Typical thermocouple and extension leads

The voltage across the thermocouple increases as the temperature rises and a suitably calibrated instrument, capable of measuring small voltages, can be used to measure the change.

The relationship between mill-volts and temperature is not linear. In microprocessor based equipment, the conversion is done based on the data stored in the device.

11.7. Resistance Temperature Detectors (RTDs)

RTD's are built from selected metals (typically Platinum), which change resistance with temperature change. Figure 11.7 shows typical RTD and thermowell construction. The transducer is the temperature sensitive resistor itself, with the sensor being a combination of the transducer and electronics that measure the resistance of the device.

The resistance temperature detector (RTD) measures the electrical conductivity as it varies with temperature. The electrical resistance generally increases with temperature, and the device is defined as having a positive temperature coefficient. The magnitude of the temperature coefficient determines the sensitivity of the RTD.

The temperature coefficient defines how much the resistance will change for a change in temperature, and has units of ohms/^{o}C. The greater the temperature coefficient, the more the resistance will change for a given change in temperature.

Apart from Platinum, other metals are used for RTDs such as Copper and Nickel. Platinum is the most common and has the best linear characteristics of the three, although Nickel has a higher temperature coefficient giving it greater sensitivity.

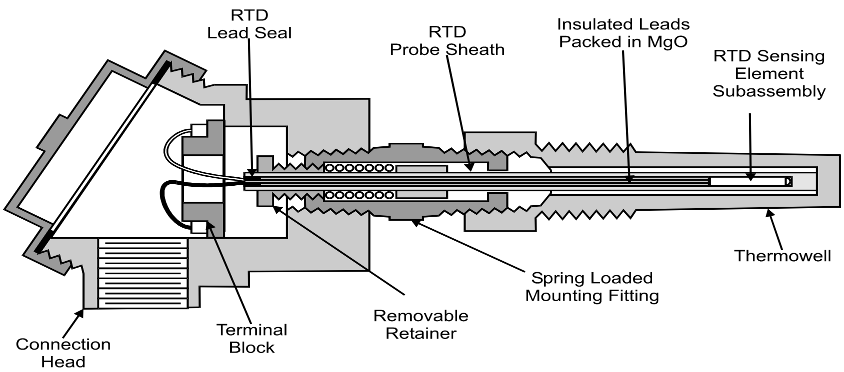


Figure 11.7
Typical RTD and thermowell construction

11.8. Thermistors

A thermistor is a semiconductor device formed from metal oxides. The principle of temperature measurement with a thermistor is that its resistance changes with temperature. Most thermistors differ from normal resistors in that they have a negative coefficient of resistance, this means that the resistance decreases with an increase in temperature.

A thermistor is a bulk semiconductor device, and as such can be fabricated in many forms. The more common include discs, beads and rods. Size does vary from a bead of 1mm to a disc of several centimeters in diameter and thickness. Wide ranges of thermistors (both resistance and temperature) are supplied by manufacturers. This is done by varying the doping and semiconductor materials.

Thermistors are not linear, and their response curves vary for the different types. Some thermistors have a near linear temperature resistance relationship, others are available with a sharp change in slope (sensitivity) at a particular characteristic temperature.

11.9. Infrared Pyrometers

Any object with a temperature above absolute zero will radiate electromagnetic energy. Infrared pyrometers measure the amount of energy radiated from an object in order to determine its temperature.

Infrared or radiation pyrometers use an optical system to focus the radiated energy onto a sensing device.

Selection and Sizing: Different materials transmit radiation of different wavelengths. In selecting an effective radiation pyrometer, a wavelength-band needs to be chosen that is not transmitted by the material. Also the selected wavelength-band cannot be absorbed by the environment.

There are a number of different types of infrared pyrometers:

- Total radiation
- Single wavelength
- Dual wavelength

11.10. Acoustic Pyrometers

Acoustic pyrometers work on the principle that the speed of sound in a gas is dependant on the nature of the gas and its temperature. The time of flight is used, and since the distance between points is known it is possible to measure any change in conditions. This principle is adapted for liquid and solid temperature measurement also.

Typical Applications: Acoustic pyrometers are used when requiring an average temperature or the temperature over a large area or volume of gas.

Acoustic pyrometers are useful for measuring gas temperatures inside kilns and furnaces. They work over a very large temperature range and are useful for mapping thermal contours. Unfortunately cost is often prohibitive.

11.11. Flow Measurement

11.11.1. Principles of Flow Measurement

There are four main types of liquid flow as shown in Table 11.1.

Table 11. 1
Main types of liquid flow

Type of flow	Notable Characteristics	Process material
General	Thin, clean liquids	Water, light oils and solvents
Two-phase flow	Liquids with bubbles	Beer, wet steam, unrefined petroleum
Slurry	Dirty liquids	Water and sand
Non-Newtonian	Heavy thick liquids	Grease, Paint, honey

Flow Measurements : There are three main measurements that can be made with process flows:

- Velocity
- Volumetric flow
- Mass flow

Velocity is the speed at which the fluid moves. This by itself does not give any information about the quantity of fluid.

Volumetric flow is often deduced by knowing the cross sectional area of the fluid. Most volumetric flow equipment measures the velocity and calculates the volumetric flow based on a constant cross sectional area.

$$Q = v \cdot A$$

where: v is the velocity
A is the cross sectional area
Q is the volumetric flow rate

Mass flow rate can only be calculated from the velocity or the volumetric flow rates if the density is constant. If the density is not constant, then mass flow measuring equipment is required for mass flow rate.

$$W = Q \cdot \rho$$

Where: Q is the volumetric flow rate
ρ is the density of the fluid
W is the mass flow rate

The flow of gases is normally measured in terms of mass per unit time. While most liquids are nearly incompressible, densities of gases vary with operating temperature and pressure. Some flowmeters, such as Coriolis meters, measure the mass flow directly. Volumetric flowmeters do not measure mass flow directly. Mass flow is calculated from the density and the volumetric flow as shown above. Some volumetric meters infer density based on the measured pressure and temperature of the fluid. This type of measurement is referred to as the inferred method of measuring mass flow as shown in Figure 11.8.

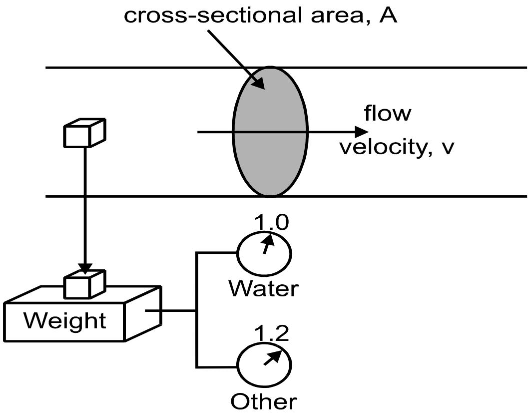


Figure 11. 8
Inferred method of measuring mass flow

11.12. Differential Pressure Flowmeters

One of the most common methods of measuring flow is with a differential flowmeter. This technique requires the pressure to be measured on both sides of an imposed restriction in the path of normal flow. The flow rate of the material can be calculated from the change in pressure.

Differential pressure devices works on the principle of inducing a change in pressure by placing a restriction in the line of flow. This effectively changes some potential energy to kinetic energy - this is detected and measured as a change in pressure.

The restriction in the pipe is called the primary element. The secondary element is the differential pressure sensor and transmitter. This device can measure and calculate the flow rate. Differential pressure or flow rate information can then be accessed from this device.

The velocity of flow is related to the square root of the differential pressure.

When the pressure differential is measured, the volumetric or mass flow rate can be calculated based on:

- fluid properties
- cross-sectional area
- shape of restriction
- adjacent piping

Formulae: The relationship between the flow rate and the change in pressure can be shown as:

$$V = k.\sqrt{h/\rho}$$

$$Q = k.A.\sqrt{h/\rho}$$

$$W = k.A.\sqrt{h\rho}$$

where: v is the velocity

Q is the volumetric flow rate
W is the mass flow rate
k is a constant
h is the differential pressure
ρ is the density of the fluid
A is the cross sectional area of the pipe

11.12.1. Open Channel Flow Measurement

The primary devices used in open channel flow measurements are weirs and flumes.

Weirs: Weirs are openings in the top of a dam or reservoir that allow for the flow of liquid and enable measurement of the flow as shown in Figure 11. 9. With the characteristics of the weir known, the flow is generally determined by the height of the liquid in the weir.

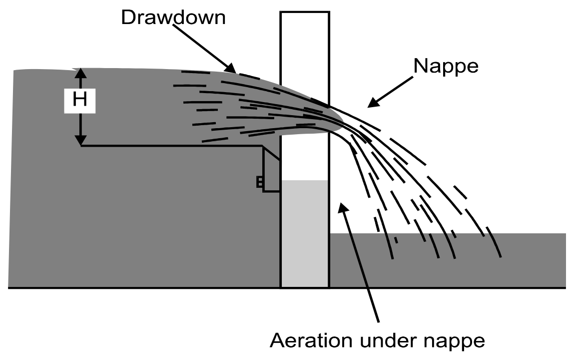


Figure 11. 9
Flow over a weir

Flumes: Flumes are a modification to the weir where the section of flow is reduced to maintain head pressure. A flume forces the liquid into a narrower channel and in doing so only incurs a head pressure drop of about 1/4 of that for a weir of equal size.

Level Measurement: The head used for calculating the flow rate is measured using level sensing equipment. Sme of the more common types for open channel flow measurement are:

- Float and cable
- Ball float
- Air bubbles
- Pressure sensing

In closed applications, and even in pipes, ultrasonics is often used to determine the level.

11.12.2. Variable Area Flowmeters

Variable area flow meters work with low viscous liquids at high velocities. The principle of operation is that the flow stream displaces a float placed in the stream.

The rate of flow is related to the area produced by forcing the float up or down, and varying the area.

It is because of the low viscosity and high velocity that the frictional resistance of the flow is negligible compared to the resistance of the obstruction (float) placed in the flow stream.

The float in the early stages of development was slotted which caused the floats to rotate. This provided stability and centering of the float, and is where the designation of rotameter came from.

The rotameter consists of a tapered measuring tube and a float. This arrangement produces a resistance value (coefficient of resistance) for the float, which depends on its position in the measuring tube. A balance is achieved between the force of the flow stream and the weight of the float. The float positions itself vertically within the measuring tube such that the resistance value is balanced. Figure 11.10 shows variable area (rotameter) flowmeter.

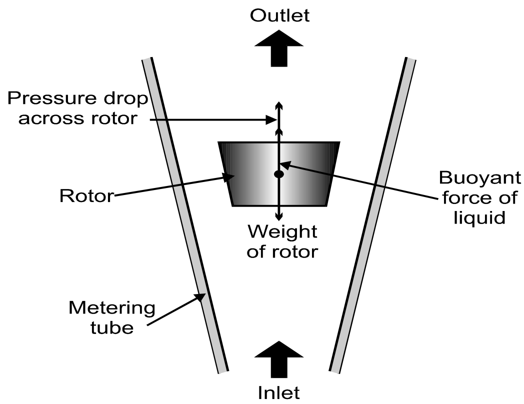


Figure 11.10
Variable area (Rotameter) flowmeter

11.13. Magnetic Flowmeters

Electromagnetic flowmeters, also known as magmeters, use Faraday's law of electromagnetic induction to sense the velocity of fluid flow.

Faraday's law states that moving a conductive material at right angles through a magnetic field induces a voltage proportional to the velocity of the conductive material. The conductive material in the case of a magmeter is the conductive fluid.

11.13.1. Ultrasonic Flow Measuremt

There are two types of ultrasonic flow measurement:

- Transit time measurement
- Doppler effect

The fundamental difference is that the transit-time method should be used for clean fluids, while the Doppler reflection type should be used for dirty, slurry type flows.

Transit Time: The transit-time flowmeter device sends pulses of ultrasonic energy diagonally across the pipe. The transit-time is measured from when the transmitter sends the pulse to when the receiver detects the pulse. Figure 11.11 shows transit time measurement.

Each location contains a transmitter and receiver. The pulses are sent alternatively upstream and downstream and the velocity of the flow is calculated from the time difference between the two directions.

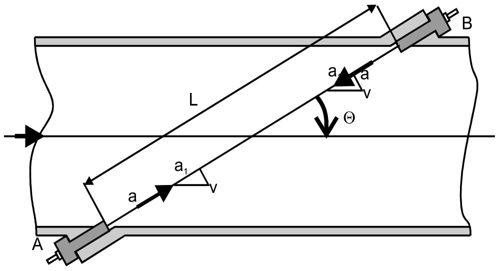


Figure 11.11
Transit time measurement

Doppler Effect: The Doppler effect device relies on objects with varying density in the flowstream to return the ultrasonic energy. With the Doppler Effect meter, a beam of ultrasonic energy is transmitted diagonally through the pipe. Portions of this ultrasonic energy are reflected back from particles in the stream of varying density. Since the objects are moving, the reflected ultrasonic energy has a different frequency. The amount of difference between the original and returned signals is proportional to the flow velocity. Figure 11.12 shows the Doppler Effect.

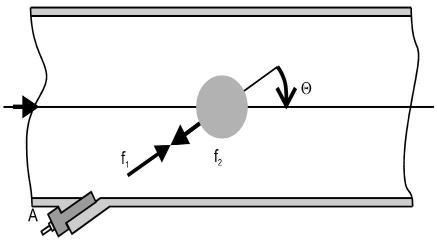


Figure 11.12
Doppler effect

It is quite common for only one sensor to be used. This contains both the transmitter and receiver. These can also be mounted outside of the pipe.

11.13.2. Mass Flow Meters

Mass flow measurement gives a more accurate account of fluids, and is not affected by density, pressure and temperature.

Although most meters can infer mass flow rate from volumetric flow measurements, there are a number of ways to measure mass flow directly:

- The Coriolis meter
- The thermal mass flowmeter
- Radiation density

11.14. Control Valves

A Control Valve is a power-operated device used to modify the fluid flow rate in a process system. Process plants consist of hundreds, or even thousands, of control circuits all networked together to produce a product. Each of these control circuits or loops is designed to maintain the plant in safe operating limits. Each of these control loops is designed to keep some important process variable such as:

- Pressure
- Flow
- Level
- Temperature

The control valve assembly typically consists mainly of the valve body, the internal trim parts, and an actuator to provide the motive power to operate the valve. A variety of additional valve accessories, which can include positioners, transducers, supply pressure regulators, manual operators, snubbers, or limit switches are also included to add functionality of various levels and to characterize the particular valve.

Basic Types of control valves: There are two basic types of control valves: rotary and linear. Linear-motion control valves commonly have globe, gate, diaphragm, or pinch - type closures. Rotary-motion valves have ball, butterfly, or plug closures. Each type of valve has its special generic features, which may, in a given application, be either an advantage or a disadvantage.

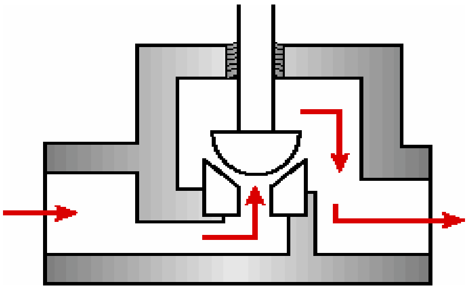


Figure 11.13
Linear valve

Linear Valve Features:

- A tortuous flow path
- A low recovery
- The throttling of small flow rates
- A variety of special trim designs
- Suitability to high-pressure applications
- Be flanged or threaded
- A separable bonnet

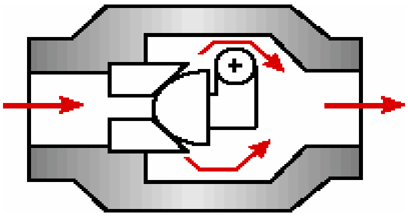


Figure 11.14
Rotary valves

Rotary Valve Features:

- Streamlined flow path
- High recovery
- More capacity
- Less packing wear
- Can handle slurry and abrasives
- Flangeless
- Integral bonnet
- High rangeability

In addition to linear and rotary, control valves are also classified according to their guiding systems and the types of services they are used in.

11.14.1. Different Types of Control Valves

The following types of control valves are commonly used in the Domestic and Engineering applications.

- Globe Valves
- Gate Valves
- Butter Fly Valves
- Eccentric Disc Valves
- Ball Valves
- Rotary Plug Valves
- Diaphragm Valves
- Pinch Valves

11.14.2. Control Valve characteristics

The performance and behavior of the control valve is dependant on the type of actuation. It also depends upon direction and the flow passage restriction. Different combinations of these parameters give a peculiar pattern of operation through the total travel of the valve opening. The flow through valve plotted against the percentage valve opening is defined as characteristic of the control valve.

The relationship between control valve flow capacity and valve stem travel is known as the flow characteristic of the control valve.

Chapter 12. Forecasts and Predictions

12.1. Main Technology Trends

Some of the main technology trends in industrial automation are discussed in this section.

Technology Shifts: The industrial automation market periodically undergoes major shifts as new technologies improve the functionality and economics of industrial monitoring and control systems. The introduction of wireless sensor networking (WSN) represents just such a sea change. While proprietary P2P and P2MP wireless technologies have been used in a limited fashion since the 1980s to integrate a few hard-to-wire field devices into an overall wired control system, standards-based WSN promises to dramatically expand the number of devices in a plant that can be connected wirelessly. In contrast to P2P and P2MP connections, WSN utilizes self-forming, self-healing mesh networking to enable field devices to be deployed cost effectively without the need for site surveys or specially trained field technicians to manually configure directional antennas. Emerson Process Management estimates that WSN enables cost savings of up to 90% compared to the deployment of wired field devices.

However, WSN will not be appropriate for all industrial automation applications. The capabilities need to be carefully assessed before selecting it.

Productivity: The fundamental purpose of industrial automation is to improve productivity – generate increased output with reduced costs and facilitate increased output by reducing the costs. The intrinsic value of each and every piece of automation equipment has the ability to provide increased productivity for the customer as well as to the users.

Productivity has now become a global race, an international competition between regions and nations for the single reason that it results in wealth and is the key to improvements in living standards. Increased productivity means things are made more cheaply and more quickly.

Knowledge workers: When reliable information is not readily available, it can lead to duplicated efforts between multiple business units, loss of sales or productivity, and poor decision making based on faulty information. All of these effects, from a lack of information, cost money.

In many situations, workers are unable to find the information they need because of inadequate, inaccurate, or delayed information. In today's information age, data needed to make decisions should be made available to all workers in an effortless fashion. Often workers waste time trying to find data by waiting on faxes or printed reports, searching through paper filing systems, or re-entering data from existing sources to create customized reports in a spreadsheet application. Data important to decision making should be easily accessible to workers through database driven, dynamic web portals or other centralized corporate applications.

A number of domain specific software is available to resolve this issue and bring significant value to productivity. These applications can be implemented either using off-the-shelf software or with specific customisation for the industry. Over the years, the automation engineers have kept pace with the changing technologies and software in particular to be able to define the specifications as desired.

Offshore Outsourcing: In this era of globalization, entrepreneurs are always looking ahead to beat competition. The new trend is - offshore outsourcing. The concept involves taking internal company functions and paying an outside firm to handle them, which enables the entrepreneurs to divert their full attention towards core competencies. Thus they can focus on their primary business.

Although Software outsourcing is the buzzword in the industry, outsourcing of all kind of business is now happening. Though the most visible benefit of this is cost savings, there are a number of other factors that influence the decision to outsource to an offshore partner. Some of them are discussed below:

12.2. The China Challenge

In today's global environment whoever manufactures products better, cheaper and faster, wins. Every country in the world is competing. In consumer products, China is grabbing a lot of the prizes. And they're moving strongly into high-tech.

Some of the facts about manufacturing in China today:

- It is continually increasing its manufacturing prowess
- There are significant cost advantages (beyond just labor cost)
- It involves a good, repetitive quality which facilitates productivity
- It has a worldwide market-share – 50% of cameras, 30% of air conditioners and televisions, 25% of washing machines, 20% of refrigerators
- One private Chinese company makes 40% of all microwave ovens sold in Europe
- The city of Wenzhou, Eastern China produces 70% of the world's metal cigarette lighters
- Walmart – Buys $18 billion from China, providing a direct link to the US consumer

12.3. Market Predictions

- Nano technology and nano scale assembly systems
- Machine to machine networking
- Bio-electronic devices
- Complex adaptive systems
- Wireless everything
- Fully automated factories

12.3.1. Nano Technology

The commercial interest in nanotechnology is being driven by visions of a stream of new nanotech commercial products and applications that will lead to a new industrial revolution - a revolution in which almost every industry is likely to be affected.

It will be possible to produce new materials with desired properties: smaller, stronger, tougher, lighter and more resilient than anything that has ever been made. Molecule-size components are being assembled into complex composites and "smart" materials. For example, nano structured membranes are being developed for efficient filtering of pollutants from water or air.

With nanotech, today's supercomputer could become tomorrow's wristwatch PDA.

12.3.2. Machine to machine networking

The convergence of smart devices with the internet is creating a new inflection point. Manufacturers can use their connected products to develop customer service relationships that can ultimately recreate the basis of customer management and generate new revenue streams in an information economy.

This will far surpass human communications in scope, value, and sheer numbers. Within the next few years, more machines will be connected via the internet than humans. Eventually reaching tens of billions of connections, machines will communicate with each other, as well as with data mining and processing systems that will automate the communication and interpretation of the mass of data they gather. This will add significant value for businesses and consumers.

12.3.3. Complex adaptive systems

Complex adaptive systems yield significant advances through reduced software, faster and easier installation, robust performance, vastly improved flexibility, capability to handle very much larger I/O point-counts. Traditional concepts of fault-tolerance become obsolete, because redundancy is provided directly at the I/O level. Complex adaptive systems are robust because the behavior is not dependent on single, or even multiple failure points. Failure of any single part of the system is accommodated. CAS has the ability to achieve much higher levels of performance through emergent behavior and self-organizing capabilities.

12.3.4. Wireless Connections

The connectivity infrastructure is moving very quickly to connect everyone and everything to the Internet, not only through high-speed DSL and cable-modems, but soon through wireless.

The impact on industrial controls will be significant. Connecting automation products with conventional wire beyond the confines of a typical system enclosure is still a major hindrance in the typical factory. This inevitably gave rise to what was previously called islands of automation.

Wireless mobility and information services already bring voice, entertainment, Internet access and safety services into cars and trucks. The automobile is quickly becoming the center of a complete range of connected appliances.

12.3.5. Fully automated factories

Automated factories and processes are too expensive to be rebuilt for every design change - so they have to be highly configurable and flexible. To successfully reconfigure an entire production line or process requires direct access to most of its control elements-switches, valves, motors and drives-down to a fine level of detail.

With technology available today, fully automated factories - in a truly realistic sense - are quickly becoming an accepted fact.